(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2008 (27.11.2008)

PCT

(10) International Publication Number
WO 2008/143679 A2

(51) International Patent Classification:
*C12N 9/24* (2006.01)

(21) International Application Number:
PCT/US2007/070284

(22) International Filing Date: 1 June 2007 (01.06.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/810,483 1 June 2006 (01.06.2006) US

(71) Applicants *(for all designated States except US)*: **VERENIUM CORPORATION** [US/US]; 4955 Directors Place, San Diego, CA 92121 (US). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US). **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US]; 1200 East California Blvd., Pasadena, CA 91125 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHANG, Cathy** [US/US]; 1318 Shadow Hills Drive, San Marcos, CA 92069 (US). **MATHUR, Eric, J.** [US/US]; 2654 Galicia Way, Carlsbag, CA 92009 (US). **CAYOUETTE, Michelle** [US/US]; 3812 Park Boulevard, #505, San Diego, CA 92103 (US). **ROBERTSON, Dan, E.** [US/US]; 6332 Oleander Way, San Diego, CA 92130 (US). **HUGENHOLTZ, Philip** [AU/US]; 2332 California Street, Berkeley, CA 94703 (US). **WARNECKE, Falk** [DE/US]; 2737 Woolsey Street, Berkeley, CA 94705 (US). **LEADBETTER, Jared** [US/US]; 1805 North Harding Avenue, Altadena, CA 91001 (US). **IVANOVA, Natalia** [RU/US]; 44 Long Creek Circle, Clayton, CA 94517 (US). **LUGINBUHL, Peter** [CH/US]; 4955 Directors Place, San Diego, CA 92121 (US). **HUTCHISON, Don** [US/US]; 343 Gardendale Road, Encinitas, CA 92024 (US).

(74) Agents: **EINHORN, Gregory, P.** et al.; Morrison & Foerster LLP, Suite 100, 12531 High Bluff Drive, San Diego, CA 92130-2040 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH,

*[Continued on next page]*

(54) Title: NUCLEIC ACIDS AND PROTEINS AND METHODS FOR MAKING AND USING THEM

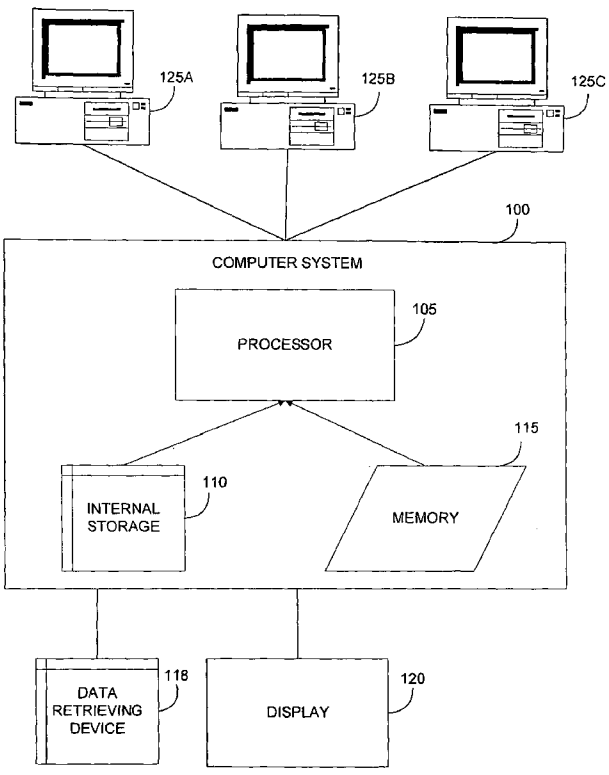



FIGURE 1

(57) Abstract: The invention provides polypeptides, including enzymes, structural proteins and binding proteins, polynucleotides encoding these polypeptides, and methods of making and using these polynucleotides and polypeptides. Polypeptides, including enzymes and antibodies, and nucleic acids of the invention can be used in industrial, experimental, food and feed processing, nutritional and pharmaceutical applications, e.g., for food and feed supplements, colorants, neutraceuticals, cosmetic and pharmaceutical needs. Polypeptides of the invention can be used in food processing, brewing, bath additives, alcohol production, peptide synthesis, enantioselectivity, hide preparation in the leather industry, waste management and animal degradation, silver recovery in the photographic industry, medical treatment, silk degumming, biofilm degradation, biomass conversion to a biofuel (e.g., a bioethanol, biomethanol, biopropanol or biobutanol, a biodiesel, etc.), biodefense, antimicrobial agents and disinfectants, personal care and cosmetics, biotech reagents, in corn wet milling and pharmaceuticals such as digestive aids and anti-inflammatory (anti-phlogistic) agents.

WO 2008/143679 A2

CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

# NUCLEIC ACIDS AND PROTEINS AND METHODS FOR MAKING AND USING THEM

## REFERENCE TO SEQUENCE LISTING SUBMITTED VIA EFS-WEB

This application is being filed electronically via the USPTO EFS-WEB server, as authorized and set forth in MPEP §1730 II.B.2(a)(A), and this electronic filing includes an electronically submitted sequence (SEQ ID) listing. The entire content of this sequence listing is herein expressly incorporated by reference for all purposes. The sequence listing is identified on the electronically filed .txt file as follows:

| File Name | Date of Creation | Size (bytes) |
|---|---|---|
| 564462014840seqlist.txt | June 1, 2007 | 199,447,624 bytes |

## REFERENCE TO TABLES SUBMITTED VIA EFS-WEB

This application is being filed electronically via the USPTO EFS-WEB server, as authorized and set forth in MPEP §1730 II.B.2(a)(A), and this electronic filing includes electronically submitted tables. The entire content of the tables is herein expressly incorporated by reference for all purposes. The tables are identified on the electronically filed .pdf files as follows:

| File Name | Date of Creation | Size (bytes) |
|---|---|---|
| Table-1.pdf | June 1, 2007 | 14.7 MB |
| Table-4.pdf | June 1, 2007 | 5.74 |

## FIELD OF THE INVENTION

This invention relates to molecular and cellular biology and biochemistry. In one aspect, the invention provides polypeptides, including enzymes, structural proteins and binding proteins (e.g., ligands, receptors), polynucleotides encoding these polypeptides, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, structural proteins and binding proteins, including thermostable and thermotolerant activity, and polynucleotides encoding these enzymes, structural proteins and binding proteins and making and using these polynucleotides and polypeptides. The polypeptides of the invention can be used in a variety of pharmaceutical, agricultural and industrial contexts, including the manufacture of cosmetics and nutraceuticals.

Additionally, the polypeptides of the invention can be used in food processing, brewing, bath additives, alcohol production, peptide synthesis, enantioselectivity, hide preparation in the leather industry, waste management and animal degradation, silver recovery in the photographic industry, medical treatment, silk degumming, biofilm degradation, biomass conversion to a biofuel (e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel, etc.), biodefense, antimicrobial agents and disinfectants, personal care and cosmetics, biotech reagents, in corn wet milling and pharmaceuticals such as digestive aids and anti-inflammatory (anti-phlogistic) agents.

## BACKGROUND

The invention provides isolated, synthetic and recombinant polypeptides, including enzymes, structural proteins and binding proteins, polynucleotides encoding these polypeptides, and methods of making and using these polynucleotides and polypeptides. The polypeptides of the invention, and the polynucleotides encoding the polypeptides of the invention, encompass many classes of enzymes, structural proteins and binding proteins. In one aspect, the enzymes and proteins of the invention include, e.g. aldolases, alpha-galactosidases, amidases, e.g. secondary amidases, amylases, catalases, carotenoid pathway enzymes, dehalogenases, endoglucanases, epoxide hydrolases, esterases, hydrolases, glucosidases, glycosidases, inteins, isomerases, laccases, lipases, monooxygenases, nitroreductases, nitrilases, P450 enzymes, pectate lyases, phosphatases, phospholipases, phytases, polymerases and xylanases. The invention also provides isolated, synthetic and recombinant polypeptides, including enzymes, structural proteins and binding proteins, polynucleotides encoding these polypeptides, having the activities described in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2. The enzymes and proteins of the invention have utility in a variety of applications.

## SUMMARY

The invention provides isolated, synthetic or recombinant nucleic acids comprising a nucleic acid sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to an exemplary nucleic acid of the

invention, e.g., including SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, etc. and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, over a region of at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500, or more residues, or over the full length of a protein coding sequence (transcript) or gene, encodes at least one polypeptide having an enzyme, structural or binding activity, and the sequence identities are determined by analysis with a sequence comparison algorithm or by a visual inspection. In one aspect, the enzymes and proteins of the invention include, e.g. aldolases, alpha-galactosidases, amidases, e.g. secondary amidases, amylases, catalases, carotenoid pathway enzymes, dehalogenases, endoglucanases, epoxide hydrolases, esterases, hydrolases, glucosidases, glycosidases, inteins, isomerases, laccases, lipases, monooxygenases, nitroreductases, nitrilases, P450 enzymes, pectate lyases, phosphatases, phospholipases, phytases, polymerases, proteases, peptidases and xylanases. In another aspect, the isolated, synthetic and recombinant polypeptides of the invention, including enzymes, structural proteins and binding proteins, and polynucleotides encoding these polypeptides, of the invention have activity as described in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

In one aspect, the invention also provides isolated, synthetic or recombinant nucleic acids with a common novelty in that they are all derived from a common source, e.g., an environmental source, mixed environmental sources or mixed cultures. The invention provides isolated, synthetic or recombinant nucleic acids isolated from a common source, e.g. an environmental source, mixed environmental sources or mixed cultures comprising a polynucleotide of the invention, e.g., an exemplary sequence of the invention, including SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, etc. and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699 over a region of at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750,

800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500, or more residues, or over the full length of a protein coding sequence (transcript) or gene, encodes at least one polypeptide having an enzyme, structural or binding activity, and the sequence identities are determined by analysis with a sequence comparison algorithm or by a visual inspection. In one aspect, the enzymes and proteins of the invention include, e.g. aldolases, alpha-galactosidases, amidases, e.g. secondary amidases, amylases, catalases, carotenoid pathway enzymes, dehalogenases, endoglucanases, epoxide hydrolases, esterases, hydrolases, glucosidases, glycosidases, inteins, isomerases, laccases, lipases, monooxygenases, nitroreductases, nitrilases, P450 enzymes, pectate lyases, phosphatases, phospholipases, phytases, polymerases, proteases, peptidases and xylanases. In another aspect, the isolated, synthetic and recombinant polypeptides of the invention, including enzymes, structural proteins and binding proteins, and polynucleotides encoding these polypeptides, of the invention have activity as described in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

In alternative aspects, the isolated, synthetic or recombinant nucleic acid encodes a polypeptide comprising an exemplary sequence of the invention, e.g., including sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, etc. and all polypeptides (amino acid sequences) disclosed in the SEQ ID listing, which include all even numbered (amino acid) SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699, and enzymatically active fragments thereof. In one aspect these polypeptides have an enzyme, structural or binding activity. In one aspect, the enzymes and proteins of the invention include, e.g. aldolases, alpha-galactosidases, amidases, e.g. secondary amidases, amylases, catalases, carotenoid pathway enzymes, dehalogenases, endoglucanases, epoxide hydrolases, esterases, hydrolases, glucosidases, glycosidases, inteins, isomerases, laccases, lipases, monooxygenases, nitroreductases, nitrilases, P450 enzymes, pectate lyases, phosphatases, phospholipases, phytases, polymerases, proteases, peptidases and/or xylanases. In another aspect, the isolated, synthetic and recombinant polypeptides of the invention, including enzymes, structural proteins and binding proteins, and polynucleotides encoding these polypeptides, of the invention have activity as described in Table 1, Table

3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

In alternative embodiments, nucleic acids (polynucleotides) of the invention encode a polypeptide of the invention having at least one conservative amino acid substitution and retaining its enzyme activity, binding activity and/or structural activity; wherein the at least one conservative amino acid substitution comprises substituting an amino acid with another amino acid of like characteristics; or, a conservative substitution comprises: replacement of an aliphatic amino acid with another aliphatic amino acid; replacement of a Serine with a Threonine or vice versa; replacement of an acidic residue with another acidic residue; replacement of a residue bearing an amide group with another residue bearing an amide group; exchange of a basic residue with another basic residue; or replacement of an aromatic residue with another aromatic residue.

In alternative embodiments, the nucleic acid (polynucleotide) of the invention encodes a polypeptide of the invention having an enzyme activity, binding activity and/or structural activity, but lacking a signal sequence, a prepro domain, a dockerin domain, and/or a carbohydrate binding module (CBM), wherein optionally the carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module.

In alternative embodiments, the nucleic acid (polynucleotide) of the invention encodes a polypeptide further comprising a heterologous sequence, wherein the heterologous sequence can comprise, or consist of, a sequence encoding: (i) a heterologous signal sequence, a heterologous carbohydrate binding module, a heterologous dockerin domain, a heterologous catalytic domain (CD), or a combination thereof; (ii) the sequence of (ii), wherein the heterologous signal sequence, carbohydrate binding module or catalytic domain (CD) is derived from a heterologous enzyme; or, (iii) a tag, an epitope, a targeting peptide, a cleavable sequence, a detectable moiety or an enzyme; or the heterologous carbohydrate binding module (CBM) can comprise, or consist of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module; or alternatively the heterologous signal sequence can target the encoded protein to a vacuole, the endoplasmic reticulum, a chloroplast or a starch granule.

In alternative aspects, the enzyme, structural or binding activity comprises a recombinase activity, a helicase activity, a DNA replication activity, a DNA

recombination activity, an isomerase, a trans-isomerase activity or topoisomerase activity, a methyl transferase activity, an aminotransferase activity, a uracil-5- methyl transferase activity, a cysteinyl tRNA synthetase activity, a hydrolase, an esterase activity, a phosphoesterase activity, an acetylmuramyl pentapeptide phosphotransferase activity, a glycosyltransferase activity, an acetyltransferase activity, an acetylglucosamine phosphate transferase activity, a centromere binding activity, a telomerase activity or a transcriptional regulatory activity, a heat shock protein activity, a protease activity, a proteinase activity, a peptidase activity, a carboxypeptidase activity, an endonuclease activity, an exonuclease activity, a RecB family exonuclease activity, a polymerase activity, a carbamoyl phosphate synthetase activity, a methyl-thioadenine synthetase activity, an oxidoreductase activity, an Fe-S oxidoreductase activity, a flavodoxin reductase activity, a permease activity, a thymidylate activity, a dehydrogenase activity, a pyrophosphorylase activity, a coenzyme metabolism activity, a dinucleotide-utilizing enzyme activity, a molybdopterin or thiamine biosynthesis activity, a beta-lactamase activity, a ligand binding activity, an ion transport activity, an ion metabolism activity, a tellurite resistance protein activity, an inorganic ion transport activity, a nucleotide transport activity, a nucleotide metabolism activity, an actin or myosin activity, a lipase activity or a lipid acyl hydrolase (LAH) activity, a cell envelop biogenesis activity, an outer membrane synthesis activity, a ribosomal structure synthesis activity, a translational processing activity, a transcriptional initiation activity, a TATA-binding activity, a signal transduction activity, an energy metabolism activity, an ATPase activity, an information storage and/or processing activity, and/or any of the polypeptides activities as set forth in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

In one aspect, the sequence comparison algorithm is a BLAST version 2.2.2 algorithm where a filtering setting is set to blastall -p blastp -d "nr pataa" -F F, and all other options are set to default.

Another aspect of the invention is an isolated, synthetic or recombinant nucleic acid including at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500, or more consecutive bases of a nucleic acid sequence of the invention, sequences substantially identical thereto, and the sequences complementary thereto.

In one aspect, the isolated, synthetic or recombinant nucleic acid encodes a polypeptide having a enzyme, structural or binding activity, that is thermostable. The thermostable polypeptide according to the invention can retain activity under conditions comprising a temperature range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 37°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. The thermostable polypeptides according to the invention can retain activity, in temperatures in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. In some embodiments, the thermostable polypeptides according to the invention retains an aldolase activity at a temperature in the ranges described above, at about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about pH 8.5, about pH 9.0, about pH 9.5, about pH 10.0, about pH 10.5, about pH 11.0, about pH 11.5, about pH 12.0 or more.

In another aspect, the isolated, synthetic or recombinant nucleic acid encodes a polypeptide having an enzyme, structural or binding activity, which is thermotolerant. The thermotolerant polypeptides according to the invention can retain activity after exposure to conditions comprising a temperature in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about

90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. The thermotolerant polypeptides according to the invention can retain activity, after exposure to a temperature in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. In some embodiments, the thermotolerant polypeptides according to the invention retains an aldolase activity after exposure to a temperature in the ranges described above, at about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about pH 8.5, about pH 9.0, about pH 9.5, about pH 10.0, about pH 10.5, about pH 11.0, about pH 11.5, about pH 12.0 or more.

The invention provides isolated, synthetic or recombinant nucleic acids comprising a sequence that hybridizes under stringent conditions to a nucleic acid comprising a sequence of the invention, e.g., an exemplary sequence of the invention, including SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, or fragments or subsequences thereof. In one aspect, the nucleic acid encodes a polypeptide having a enzyme, structural or binding activity. The nucleic acid can be at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200 or more residues in length or the full length of the gene or transcript. In one aspect, the stringent conditions include a wash step comprising a wash in 0.2X SSC at a temperature of about 65°C for about 15 minutes.

The invention provides a nucleic acid probe for identifying a nucleic acid encoding a polypeptide having a enzyme, structural or binding activity, wherein the probe comprises at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 or more, consecutive bases of a sequence comprising a sequence of the invention, or fragments or subsequences thereof, wherein the probe identifies the nucleic acid by binding or hybridization. The probes of this invention can comprise an oligonucleotide comprising between about 10 to 100 consecutive bases of a sequence in accordance with the invention, or fragments or subsequences thereof, for example, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 bases or more, or, any desired length in between. The probe can comprise an oligonucleotide comprising at least about 10 to 50, about 20 to 60, about 30 to 70, about 40 to 80, or about 60 to 100 consecutive bases of a sequence comprising a sequence of the invention, or fragments or subsequences thereof.

The invention provides a nucleic acid probe for identifying a nucleic acid encoding a polypeptide having a enzyme, structural or binding activity, wherein the probe comprises a nucleic acid comprising a sequence at least about 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 or more residues having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to a nucleic acid of the invention. In one aspect, the sequence identities are determined by analysis with a sequence comparison algorithm or by visual inspection. In other embodiments, the probes can comprise an oligonucleotide comprising between at least about 10 to 100 consecutive bases of a nucleic acid sequence in accordance with the invention, or a subsequence thereof, for example about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 bases or more, or, any desired length in between.

The invention provides an amplification primer pair for amplifying a nucleic acid encoding a polypeptide having a enzyme, structural or binding activity, wherein the primer pair is capable of amplifying a nucleic acid comprising a sequence of the invention, or fragments or subsequences thereof. One or each member of the amplification primer sequence pair can comprise an oligonucleotide comprising at least

about 10 to 50, or more, consecutive bases of the sequence, or about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more consecutive bases of the sequence.

The invention provides amplification primer pairs, wherein the primer pair comprises a first member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 or more residues of a nucleic acid of the invention, and a second member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 or more residues of the complementary strand of the first member.

The invention provides polypeptide-, enzyme-, protein-, e.g. structural or binding protein-encoding nucleic acids generated by amplification, e.g., polymerase chain reaction (PCR), using an amplification primer pair of the invention. The invention provides polypeptide-, enzyme-, protein-, e.g. structural or binding protein-encoding nucleic acids generated by amplification, e.g., polymerase chain reaction (PCR), using an amplification primer pair of the invention. The invention provides methods of making a polypeptide, enzyme, protein, e.g. structural or binding protein, by amplification, e.g., polymerase chain reaction (PCR), using an amplification primer pair of the invention. In one aspect, the amplification primer pair amplifies a nucleic acid from a library, e.g., a gene library, such as an environmental library.

The invention provides methods of amplifying a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity, comprising amplification of a template nucleic acid with an amplification primer sequence pair capable of amplifying a nucleic acid sequence of the invention, or fragments or subsequences thereof.

The invention provides expression cassettes comprising a nucleic acid of the invention or a subsequence thereof. In one aspect, the expression cassette can comprise the nucleic acid that is operably linked to a promoter. The promoter can be a viral, bacterial, mammalian or plant promoter. In one aspect, the plant promoter can be a potato, rice, corn, wheat, tobacco or barley promoter. The promoter can be a constitutive promoter. The constitutive promoter can comprise CaMV35S. In another aspect, the promoter can be an inducible promoter. In one aspect, the promoter can be a tissue-specific promoter or an environmentally regulated or a developmentally regulated promoter. Thus, the promoter can be, e.g., a seed-specific, a leaf-specific, a root-specific,

a stem-specific or an abscission-induced promoter. In one aspect, the expression cassette can further comprise a plant or plant virus expression vector.

The invention provides cloning vehicles comprising an expression cassette (e.g., a vector) of the invention or a nucleic acid of the invention. The cloning vehicle can be a viral vector, a plasmid, a phage, a phagemid, a cosmid, a fosmid, a bacteriophage or an artificial chromosome. The viral vector can comprise an adenovirus vector, a retroviral vector or an adeno-associated viral vector. The cloning vehicle can comprise a bacterial artificial chromosome (BAC), a plasmid, a bacteriophage P1-derived vector (PAC), a yeast artificial chromosome (YAC), or a mammalian artificial chromosome (MAC).

The invention provides transformed cell comprising a nucleic acid of the invention or an expression cassette (e.g., a vector) of the invention, or a cloning vehicle of the invention. In one aspect, the transformed cell can be a bacterial cell, a mammalian cell, a fungal cell, a yeast cell, an insect cell or a plant cell. In one aspect, the plant cell can be a cereal, a potato, wheat, rice, corn, tobacco or barley cell.

The invention provides transgenic non-human animals comprising a nucleic acid of the invention or an expression cassette (e.g., a vector) of the invention. In one aspect, the animal is a mouse, a rat, a pig, a goat or a sheep.

The invention provides transgenic plants comprising a nucleic acid of the invention or an expression cassette (e.g., a vector) of the invention. The transgenic plant can be a cereal plant, a corn plant, a potato plant, a tomato plant, a wheat plant, an oilseed plant, a rapeseed plant, a soybean plant, a rice plant, a barley plant or a tobacco plant.

The invention provides transgenic seeds comprising a nucleic acid of the invention or an expression cassette (e.g., a vector) of the invention. The transgenic seed can be a cereal plant, a corn seed, a wheat kernel, an oilseed, a rapeseed, a soybean seed, a palm kernel, a sunflower seed, a sesame seed, a peanut or a tobacco plant seed.

The invention provides an antisense oligonucleotide comprising a nucleic acid sequence complementary to or capable of hybridizing under stringent conditions to a nucleic acid of the invention. The invention provides methods of inhibiting the translation of a polypeptide, enzyme, protein, e.g. structural or binding protein message in a cell comprising administering to the cell or expressing in the cell an antisense oligonucleotide comprising a nucleic acid sequence complementary to or capable of hybridizing under stringent conditions to a nucleic acid of the invention. In one aspect, the antisense oligonucleotide is between about 10 to 50, about 20 to 60, about 30 to 70,

about 40 to 80, or about 60 to 100 bases in length, e.g., 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more bases in length.

The invention provides methods of inhibiting the translation of a polypeptide, enzyme, protein, e.g. structural or binding protein message in a cell comprising administering to the cell or expressing in the cell an antisense oligonucleotide comprising a nucleic acid sequence complementary to or capable of hybridizing under stringent conditions to a nucleic acid of the invention. The invention provides double-stranded inhibitory RNA (RNAi, or RNA interference) molecules (including small interfering RNA, or siRNAs, for inhibiting transcription, and microRNAs, or miRNAs, for inhibiting translation) comprising a subsequence of a sequence of the invention. In one aspect, the RNAi is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more duplex nucleotides in length. The invention provides methods of inhibiting the expression of a polypeptide, enzyme, protein, peptide, e.g. structural or binding protein in a cell comprising administering to the cell or expressing in the cell a double-stranded inhibitory RNA (iRNA, including small interfering RNA, or siRNAs, for inhibiting transcription, and microRNAs, or miRNAs, for inhibiting translation), wherein the RNA comprises a subsequence of a sequence of the invention.

The invention provides isolated, synthetic or recombinant polypeptides encoded by a nucleic acid of the invention, and enzymatically active fragments thereof. In alternative aspects, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides (amino acid sequences) disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699 (the exemplary sequences of the invention), or subsequences thereof, including fragments having enzymatic and/or substrate binding activity. The polypeptide can have an enzyme, structural or binding activity.

In alternative aspects, the enzyme, structural or binding activity comprises a recombinase activity, a helicase activity, a DNA replication activity, a DNA recombination activity, an isomerase, a trans-isomerase activity or topoisomerase activity, a methyl transferase activity, an aminotransferase activity, a uracil-5- methyl transferase activity, a cysteinyl tRNA synthetase activity, a hydrolase, an esterase activity, a phosphoesterase activity, an acetylmuramyl pentapeptide phosphotransferase activity, a glycosyltransferase activity, an acetyltransferase activity, an acetylglucosamine phosphate

transferase activity, a centromere binding activity, a telomerase activity or a transcriptional regulatory activity, a heat shock protein activity, a protease activity, a proteinase activity, a peptidase activity, a carboxypeptidase activity, an endonuclease activity, an exonuclease activity, a RecB family exonuclease activity, a polymerase activity, a carbamoyl phosphate synthetase activity, a methyl-thioadenine synthetase activity, an oxidoreductase activity, an Fe-S oxidoreductase activity, a flavodoxin reductase activity, a permease activity, a thymidylate activity, a dehydrogenase activity, a pyrophosphorylase activity, a coenzyme metabolism activity, a dinucleotide-utilizing enzyme activity, a molybdopterin or thiamine biosynthesis activity, a beta-lactamase activity, a ligand binding activity, an ion transport activity, an ion metabolism activity, a tellurite resistance protein activity, an inorganic ion transport activity, a nucleotide transport activity, a nucleotide metabolism activity, an actin or myosin activity, a lipase activity or a lipid acyl hydrolase (LAH) activity, a cell envelop biogenesis activity, an outer membrane synthesis activity, a ribosomal structure synthesis activity, a translational processing activity, a transcriptional initiation activity, a TATA-binding activity, a signal transduction activity, an energy metabolism activity, an ATPase activity, an information storage and/or processing activity, and/or any of the polypeptide activities as set forth in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

Exemplary polypeptide or peptide (amino acid) sequences of the invention include SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699, and subsequences thereof and variants thereof. Exemplary polypeptides also include fragments of at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or more residues in length, or over the full length of an enzyme. Exemplary polypeptide or peptide sequences of the invention include sequence encoded by a nucleic acid of the invention. Exemplary polypeptide or peptide sequences of the invention include polypeptides or peptides specifically bound by an antibody of the invention.

In alternative embodiments, a polypeptide of the invention can have at least one conservative amino acid substitution and retain its enzyme activity, binding activity and/or structural activity; wherein the at least one conservative amino acid substitution comprises substituting an amino acid with another amino acid of like characteristics; or, a conservative substitution comprises: replacement of an aliphatic amino acid with another aliphatic amino

acid; replacement of a Serine with a Threonine or vice versa; replacement of an acidic residue with another acidic residue; replacement of a residue bearing an amide group with another residue bearing an amide group; exchange of a basic residue with another basic residue; or replacement of an aromatic residue with another aromatic residue.

In alternative embodiments, a polypeptide of the invention can have an enzyme activity, binding activity and/or structural activity but lacking a signal sequence, a prepro domain, a dockerin domain, and/or a carbohydrate binding module (CBM), wherein optionally the carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module.

In alternative embodiments, a polypeptide can further comprise a heterologous sequence, wherein the heterologous sequence can comprise, or consist of, a sequence encoding: (i) a heterologous signal sequence, a heterologous carbohydrate binding module, a heterologous dockerin domain, a heterologous catalytic domain (CD), or a combination thereof; (ii) the sequence of (ii), wherein the heterologous signal sequence, carbohydrate binding module or catalytic domain (CD) is derived from a heterologous enzyme; or, (iii) a tag, an epitope, a targeting peptide, a cleavable sequence, a detectable moiety or an enzyme; or the heterologous carbohydrate binding module (CBM) can comprise, or consist of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module; or alternatively the heterologous signal sequence can target the encoded protein to a vacuole, the endoplasmic reticulum, a chloroplast or a starch granule.

In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein, is thermostable. The polypeptide, enzyme, protein, e.g. structural or binding protein can retain activity under conditions comprising a temperature range of between about 1°C to about 5°C, between about 5°C to about 15°C, between about 15°C to about 25°C, between about 25°C to about 37°C, between about 37°C to about 95°C, between about 55°C to about 85°C, between about 70°C to about 75°C, or between about 90°C to about 95°C, or more. In another aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein can be thermotolerant. The polypeptide, enzyme, protein, e.g. structural or binding protein can retain activity after exposure to a temperature in the range from greater than 37°C to about 95°C, or in the range from greater than 55°C to about 85°C. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein can retain

activity after exposure to a temperature in the range from greater than 90°C to about 95°C at pH 4.5.

Another aspect of the invention provides an isolated, synthetic or recombinant polypeptide or peptide including at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 or more consecutive bases of a polypeptide or peptide sequence of the invention, sequences substantially identical thereto, and the sequences complementary thereto. The peptide can be, e.g., an immunogenic fragment, a motif (e.g., a binding site), a signal sequence, a prepro sequence or an active site.

The invention provides isolated, synthetic or recombinant nucleic acids comprising a sequence encoding a polypeptide, enzyme, protein, e.g. structural or binding protein having any of the activities as set forth in Tables 1, 2 or 3, and a signal sequence, wherein the nucleic acid comprises a sequence of the invention. In one aspect, the isolated, synthetic or recombinant polypeptide can comprise the polypeptide of the invention comprising a heterologous signal sequence or a heterologous preprosequence, such as a heterologous enzyme or non-enzyme signal sequence. The invention provides isolated, synthetic or recombinant nucleic acids comprising a sequence encoding a polypeptide, enzyme, protein, e.g. structural or binding protein having any of the activities as set forth in Tables 1, 2 or 3, wherein the sequence does not contain a signal sequence and the nucleic acid comprises a sequence of the invention. In one aspect, the invention provides an isolated, synthetic or recombinant polypeptide comprising a polypeptide of the invention lacking all or part of a signal sequence.

In one aspect, the invention provides chimeric proteins comprising a first domain comprising a signal sequence of the invention and at least a second domain. The protein can be a fusion protein. The second domain can comprise an enzyme. The enzyme can be a non-enzyme.

The invention provides chimeric polypeptides comprising at least a first domain comprising signal peptide (SP), a prepro sequence and/or a catalytic domain (CD) of the invention and at least a second domain comprising a heterologous polypeptide or peptide, wherein the heterologous polypeptide or peptide is not naturally associated with the signal peptide (SP), prepro sequence and/ or catalytic domain (CD). In one aspect, the heterologous polypeptide or peptide is not an enzyme. The heterologous polypeptide or peptide can be amino terminal to, carboxy terminal to or on both ends of the signal peptide (SP), prepro sequence and/or catalytic domain (CD).

The invention provides isolated, synthetic or recombinant nucleic acids encoding a chimeric polypeptide, wherein the chimeric polypeptide comprises at least a first domain comprising signal peptide (SP), a prepro domain and/or a catalytic domain (CD) of the invention and at least a second domain comprising a heterologous polypeptide or peptide, wherein the heterologous polypeptide or peptide is not naturally associated with the signal peptide (SP), prepro domain and/ or catalytic domain (CD).

The invention provides isolated, synthetic or recombinant signal sequences (e.g., signal peptides) consisting of or comprising a sequence as set forth in residues 1 to 14, 1 to 15, 1 to 16, 1 to 17, 1 to 18, 1 to 19, 1 to 20, 1 to 21, 1 to 22, 1 to 23, 1 to 24, 1 to 25, 1 to 26, 1 to 27, 1 to 28, 1 to 28, 1 to 30, 1 to 31, 1 to 32, 1 to 33, 1 to 34, 1 to 35, 1 to 36, 1 to 37, 1 to 38, 1 to 40, 1 to 41, 1 to 42, 1 to 43, 1 to 44, 1 to 45, 1 to 46 or 1 to 47, of a polypeptide of the invention, including the exemplary polypeptides of the invention (including SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699). In one aspect, the invention provides signal sequences comprising the first 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 or more amino terminal residues of a polypeptide of the invention.

In one aspect, the enzyme, structural or binding activity comprises a specific activity at about 37°C in the range from about 1 to about 1200 units per milligram of protein, or, about 100 to about 1000 units per milligram of protein. In another aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein activity comprises a specific activity from about 100 to about 1000 units per milligram of protein, or, from about 500 to about 750 units per milligram of protein. Alternatively, the enzyme, structural or binding activity comprises a specific activity at 37°C in the range from about 1 to about 750 units per milligram of protein, or, from about 500 to about 1200 units per milligram of protein. In one aspect, the enzyme, structural or binding activity comprises a specific activity at 37°C in the range from about 1 to about 500 units per milligram of protein, or, from about 750 to about 1000 units per milligram of protein. In another aspect, the enzyme, structural or binding activity comprises a specific activity at 37°C in the range from about 1 to about 250 units per milligram of protein. Alternatively, the enzyme, structural or binding activity comprises a specific activity at 37°C in the range from about 1 to about 100 units per milligram of protein.

In another aspect, thermotolerance comprises retention of at least half of the specific activity of the enzyme, structural or binding protein after being heated to an elevated temperature such as a temperature from about 0°C to about 20°C, about 20°C to about 37°C, about 37°C to about 50°C, about 50°C to about 70°C, about 70°C to about 75°C, about 75°C to about 80°C, about 80°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 110°C, or higher. Alternatively, thermotolerance can comprise retention of specific activity in the range from about 1 to about 1200 units per milligram of protein, or, from about 500 to about 1000 units per milligram of protein, after being heated to an elevated temperature. In another aspect, thermotolerance can comprise retention of specific activity in the range from about 1 to about 500 units per milligram of protein after being heated to an elevated temperature, as described above.

The invention provides the isolated, synthetic or recombinant polypeptide of the invention, wherein the polypeptide comprises at least one glycosylation site. In one aspect, glycosylation can be an N-linked glycosylation. In one aspect, the polypeptide can be glycosylated after being expressed in a *P. pastoris* or a *S. pombe*.

In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein can retain activity under conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 or less (more acidic) pH. In another aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein can retain activity under conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11.0, pH 11.5, pH 12, pH 12.5 or more (more basic) pH. In one aspect, the polypeptide can retain an enzyme, structural or binding activity after exposure to conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 or less (more acidic) pH. In another aspect, the polypeptide can retain enzyme, structural or binding activity after exposure to conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11.0, pH 11.5, pH 12, pH 12.5 or more (more basic) pH.

In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein of the invention has activity at under alkaline conditions, e.g., the alkaline conditions of the gut, e.g., the small intestine. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein can retain activity after exposure to the acidic pH of the stomach.

The invention provides protein preparations comprising a polypeptide of the invention, wherein the protein preparation comprises a liquid, a solid or a gel.

The invention provides heterodimers comprising a polypeptide of the invention and a second protein or domain. The second member of the heterodimer can be a different enzyme, a different enzyme or another protein. In one aspect, the second domain can be a polypeptide and the heterodimer can be a fusion protein. In one aspect, the second domain can be an epitope or a tag. In one aspect, the invention provides homodimers comprising a polypeptide of the invention.

The invention provides immobilized polypeptides having enzyme, structural or binding activity, wherein the polypeptide comprises a polypeptide of the invention, a polypeptide encoded by a nucleic acid of the invention, or a polypeptide comprising a polypeptide of the invention and a second domain. In one aspect, the polypeptide can be immobilized on a cell, a metal, a resin, a polymer, a ceramic, a glass, a microelectrode, a graphitic particle, a bead, a gel, a plate, an array or a capillary tube.

The invention provides arrays comprising an immobilized nucleic acid of the invention. The invention provides arrays comprising an antibody of the invention.

The invention provides isolated, synthetic or recombinant antibodies that specifically bind to a polypeptide of the invention or to a polypeptide encoded by a nucleic acid of the invention. These antibodies of the invention can be a monoclonal or a polyclonal antibody. The invention provides hybridomas comprising an antibody of the invention, e.g., an antibody that specifically binds to a polypeptide of the invention or to a polypeptide encoded by a nucleic acid of the invention. The invention provides nucleic acids encoding these antibodies.

The invention provides method of isolating or identifying a polypeptide having enzyme, structural or binding activity comprising the steps of: (a) providing an antibody of the invention; (b) providing a sample comprising polypeptides; and (c) contacting the sample of step (b) with the antibody of step (a) under conditions wherein the antibody can specifically bind to the polypeptide, thereby isolating or identifying a polypeptide having an enzyme, structural or binding activity.

The invention provides methods of making an anti-polypeptide, anti-enzyme, or anti-protein, e.g. anti-structural or anti-binding protein, antibody comprising administering to a non-human animal a nucleic acid of the invention or a polypeptide of the invention or subsequences thereof in an amount sufficient to generate a humoral immune response, thereby making an anti-polypeptide, anti-enzyme, or anti-protein, e.g. anti-structural or anti-binding protein, antibody. The invention provides methods of making an anti-polypeptide, anti-enzyme, or anti-protein, e.g. anti-structural or anti-

binding protein, immune comprising administering to a non-human animal a nucleic acid of the invention or a polypeptide of the invention or subsequences thereof in an amount sufficient to generate an immune response.

The invention provides methods of producing a recombinant polypeptide comprising the steps of: (a) providing a nucleic acid of the invention operably linked to a promoter; and (b) expressing the nucleic acid of step (a) under conditions that allow expression of the polypeptide, thereby producing a recombinant polypeptide. In one aspect, the method can further comprise transforming a host cell with the nucleic acid of step (a) followed by expressing the nucleic acid of step (a), thereby producing a recombinant polypeptide in a transformed cell.

The invention provides methods for identifying a polypeptide having enzyme, structural or binding activity comprising the following steps: (a) providing a polypeptide of the invention; or a polypeptide encoded by a nucleic acid of the invention; (b) providing an enzyme, structural or binding activity substrate; and (c) contacting the polypeptide or a fragment or variant thereof of step (a) with the substrate of step (b) and detecting a decrease in the amount of substrate or an increase in the amount of a reaction product, wherein a decrease in the amount of the substrate or an increase in the amount of the reaction product detects a polypeptide having a enzyme, structural or binding activity.

The invention provides methods for identifying a polypeptide, enzyme, protein, e.g. structural or binding protein, substrate comprising the following steps: (a) providing a polypeptide of the invention; or a polypeptide encoded by a nucleic acid of the invention; (b) providing a test substrate; and (c) contacting the polypeptide of step (a) with the test substrate of step (b) and detecting a decrease in the amount of substrate or an increase in the amount of reaction product, wherein a decrease in the amount of the substrate or an increase in the amount of a reaction product identifies the test substrate as a polypeptide, enzyme, protein, e.g. structural or binding protein, substrate.

The invention provides methods of determining whether a test compound specifically binds to a polypeptide comprising the following steps: (a) expressing a nucleic acid or a vector comprising the nucleic acid under conditions permissive for translation of the nucleic acid to a polypeptide, wherein the nucleic acid comprises a nucleic acid of the invention, or, providing a polypeptide of the invention; (b) providing a test compound; (c) contacting the polypeptide with the test compound; and (d) determining whether the test compound of step (b) specifically binds to the polypeptide.

The invention provides methods for identifying a modulator of a enzyme, structural or binding activity comprising the following steps: (a) providing a polypeptide of the invention or a polypeptide encoded by a nucleic acid of the invention; (b) providing a test compound; (c) contacting the polypeptide of step (a) with the test compound of step (b) and measuring an activity of the polypeptide, enzyme, protein, e.g. structural or binding protein, wherein a change in the enzyme, structural or binding activity measured in the presence of the test compound compared to the activity in the absence of the test compound provides a determination that the test compound modulates the enzyme, structural or binding activity. In one aspect, the enzyme, structural or binding activity can be measured by providing a polypeptide, enzyme, protein, e.g. structural or binding protein, substrate and detecting a decrease in the amount of the substrate or an increase in the amount of a reaction product, or, an increase in the amount of the substrate or a decrease in the amount of a reaction product. A decrease in the amount of the substrate or an increase in the amount of the reaction product with the test compound as compared to the amount of substrate or reaction product without the test compound identifies the test compound as an activator of enzyme, structural or binding activity. An increase in the amount of the substrate or a decrease in the amount of the reaction product with the test compound as compared to the amount of substrate or reaction product without the test compound identifies the test compound as an inhibitor of enzyme, structural or binding activity.

The invention provides computer systems comprising a processor and a data storage device wherein said data storage device has stored thereon a polypeptide sequence or a nucleic acid sequence of the invention (e.g., a polypeptide encoded by a nucleic acid of the invention). In one aspect, the computer system can further comprise a sequence comparison algorithm and a data storage device having at least one reference sequence stored thereon. In another aspect, the sequence comparison algorithm comprises a computer program that indicates polymorphisms. In one aspect, the computer system can further comprise an identifier that identifies one or more features in said sequence. The invention provides computer readable media having stored thereon a polypeptide sequence or a nucleic acid sequence of the invention. The invention provides methods for identifying a feature in a sequence comprising the steps of: (a) reading the sequence using a computer program which identifies one or more features in a sequence, wherein the sequence comprises a polypeptide sequence or a nucleic acid sequence of the invention; and (b) identifying one or more features in the sequence with the computer program. The

invention provides methods for comparing a first sequence to a second sequence comprising the steps of: (a) reading the first sequence and the second sequence through use of a computer program which compares sequences, wherein the first sequence comprises a polypeptide sequence or a nucleic acid sequence of the invention; and (b) determining differences between the first sequence and the second sequence with the computer program. The step of determining differences between the first sequence and the second sequence can further comprise the step of identifying polymorphisms. In one aspect, the method can further comprise an identifier that identifies one or more features in a sequence. In another aspect, the method can comprise reading the first sequence using a computer program and identifying one or more features in the sequence.

The invention provides methods for isolating or recovering a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, from a sample, such as an environmental sample, comprising the steps of: (a) providing an amplification primer sequence pair for amplifying a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, wherein the primer pair is capable of amplifying a nucleic acid of the invention; (b) isolating a nucleic acid from the sample or treating the sample such that nucleic acid in the sample is accessible for hybridization to the amplification primer pair; and, (c) combining the nucleic acid of step (b) with the amplification primer pair of step (a) and amplifying nucleic acid from the sample, thereby isolating or recovering a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein from a sample. One or each member of the amplification primer sequence pair can comprise an oligonucleotide comprising an amplification primer sequence pair of the invention, e.g., having at least about 10 to 50 consecutive bases of a sequence of the invention.

The invention provides methods for isolating or recovering a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein from a sample, such as an environmental sample, comprising the steps of: (a) providing a polynucleotide probe comprising a nucleic acid of the invention or a subsequence thereof; (b) isolating a nucleic acid from the sample (e.g., an environmental sample) or treating the sample such that nucleic acid in the sample is accessible for hybridization to a polynucleotide probe of step (a); (c) combining the isolated nucleic acid or the treated sample of step (b) with the polynucleotide probe of step (a); and (d) isolating a nucleic acid that specifically hybridizes with the polynucleotide probe of step (a), thereby isolating or recovering a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein

from a sample. The sample (e.g., an environmental sample) can comprise a water sample, a liquid sample, a soil sample, an air sample or a biological sample. In one aspect, the biological sample can be derived from a bacterial cell, a protozoan cell, an insect cell, a yeast cell, a plant cell, a fungal cell or a mammalian cell.

The invention provides methods of generating a variant of a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity comprising the steps of: (a) providing a template nucleic acid comprising a nucleic acid of the invention; and (b) modifying, deleting or adding one or more nucleotides in the template sequence, or a combination thereof, to generate a variant of the template nucleic acid. In one aspect, the method can further comprise expressing the variant nucleic acid to generate a variant the polypeptide, enzyme, protein, e.g. structural or binding protein. The modifications, additions or deletions can be introduced by a method comprising error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR) or a combination thereof. In another aspect, the modifications, additions or deletions are introduced by a method comprising recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and a combination thereof.

In one aspect, the method can be iteratively repeated until a polypeptide, enzyme, protein, e.g. structural or binding protein having an altered or different activity or an altered or different stability from that of a polypeptide encoded by the template nucleic acid is produced. In one aspect, the variant the polypeptide, enzyme, protein, e.g. structural or binding protein is thermotolerant, and retains some activity after being exposed to an elevated temperature. In another aspect, the variant the polypeptide, enzyme, protein, e.g. structural or binding protein has increased glycosylation as compared to the polypeptide, enzyme, protein, e.g. structural or binding protein encoded by a template nucleic acid. Alternatively, the variant the polypeptide, enzyme, protein, e.g. structural or binding protein has an enzyme, structural or binding activity under a high temperature, wherein the polypeptide, enzyme, protein, e.g. structural or binding

protein encoded by the template nucleic acid is not active under the high temperature. In one aspect, the method can be iteratively repeated until a polypeptide, enzyme, protein, e.g. structural or binding protein coding sequence having an altered codon usage from that of the template nucleic acid is produced. In another aspect, the method can be iteratively repeated until a polypeptide, enzyme, protein, e.g. structural or binding protein gene having higher or lower level of message expression or stability from that of the template nucleic acid is produced.

The invention provides methods for modifying codons in a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity to increase its expression in a host cell, the method comprising the following steps: (a) providing a nucleic acid of the invention encoding a polypeptide having an enzyme, structural or binding activity; and, (b) identifying a non-preferred or a less preferred codon in the nucleic acid of step (a) and replacing it with a preferred or neutrally used codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to increase its expression in a host cell.

The invention provides methods for modifying codons in a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity; the method comprising the following steps: (a) providing a nucleic acid of the invention; and, (b) identifying a codon in the nucleic acid of step (a) and replacing it with a different codon encoding the same amino acid as the replaced codon, thereby modifying codons in a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein.

The invention provides methods for modifying codons in a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity to increase its expression in a host cell, the method comprising the following steps: (a) providing a nucleic acid of the invention encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, polypeptide; and, (b) identifying a non-preferred or a less preferred codon in the nucleic acid of step (a) and replacing it with a preferred or neutrally used codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to increase its expression in a host cell.

The invention provides methods for modifying a codon in a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity to decrease its expression in a host cell, the method comprising the following steps: (a) providing a nucleic acid of the invention; and (b) identifying at least one preferred codon in the nucleic acid of step (a) and replacing it with a non-preferred or less preferred codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in a host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to decrease its expression in a host cell. In one aspect, the host cell can be a bacterial cell, a fungal cell, an insect cell, a yeast cell, a plant cell or a mammalian cell.

The invention provides methods for producing a library of nucleic acids encoding a plurality of modified polypeptides, enzymes, proteins, e.g. structural or binding proteins, active sites or substrate binding sites, wherein the modified active sites or substrate binding sites are derived from a first nucleic acid comprising a sequence encoding a first active site or a first substrate binding site the method comprising the following steps: (a) providing a first nucleic acid encoding a first active site or first substrate binding site, wherein the first nucleic acid sequence comprises a sequence that hybridizes under stringent conditions to a nucleic acid of the invention, and the nucleic acid encodes a polypeptide, enzyme, protein, e.g. structural or binding protein, active site or a polypeptide, enzyme, protein, e.g. structural or binding protein, substrate binding site; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of active site-encoding or substrate binding site-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of modified the polypeptide, enzyme, protein, e.g. structural or binding protein, active sites or substrate binding sites. In one aspect, the method comprises mutagenizing the first nucleic acid of step (a) by a method comprising an optimized directed evolution system, Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR), error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, and a combination thereof. In another aspect, the

method comprises mutagenizing the first nucleic acid of step (a) or variants by a method comprising recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and a combination thereof.

The invention provides methods for making a small molecule comprising the steps of: (a) providing a plurality of biosynthetic enzymes capable of synthesizing or modifying a small molecule, wherein one of the enzymes comprises an enzyme encoded by a nucleic acid of the invention; (b) providing a substrate for at least one of the enzymes of step (a); and, (c) reacting the substrate of step (b) with the enzymes under conditions that facilitate a plurality of biocatalytic reactions to generate a small molecule by a series of biocatalytic reactions.

The invention provides methods for modifying a small molecule comprising the steps: (a) providing a enzyme encoded by a nucleic acid of the invention; (b) providing a small molecule; and, (c) reacting the enzyme of step (a) with the small molecule of step (b) under conditions that facilitate an enzymatic reaction catalyzed by the enzyme, thereby modifying a small molecule by an enzymatic reaction. In one aspect, the method comprises providing a plurality of small molecule substrates for the enzyme of step (a), thereby generating a library of modified small molecules produced by at least one enzymatic reaction catalyzed by the enzyme. In one aspect, the method further comprises a plurality of additional enzymes under conditions that facilitate a plurality of biocatalytic reactions by the enzymes to form a library of modified small molecules produced by the plurality of enzymatic reactions. In one aspect, the method further comprises the step of testing the library to determine if a particular modified small molecule that exhibits a desired activity is present within the library. The step of testing the library can further comprises the steps of systematically eliminating all but one of the biocatalytic reactions used to produce a portion of the plurality of the modified small molecules within the library by testing the portion of the modified small molecule for the presence or absence of the particular modified small molecule with a desired activity, and identifying at least one specific biocatalytic reaction that produces the particular modified small molecule of desired activity.

The invention provides methods for determining a functional fragment of a polypeptide, enzyme, protein, e.g. structural or binding protein, comprising the steps of: (a) providing a polypeptide, enzyme, protein, e.g. structural or binding protein, wherein the enzyme comprises a polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention, or a subsequence thereof; and (b) deleting a plurality of amino acid residues from the sequence of step (a) and testing the remaining subsequence for an enzyme, structural or binding activity, thereby determining a functional fragment of a polypeptide, enzyme, protein, e.g. structural or binding protein. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein activity is measured by providing a polypeptide, enzyme, protein, e.g. structural or binding protein, substrate and detecting a decrease in the amount of the substrate or an increase in the amount of a reaction product.

The invention provides methods for whole cell engineering of new or modified phenotypes by using real-time metabolic flux analysis, the method comprising the following steps: (a) making a modified cell by modifying the genetic composition of a cell, wherein the genetic composition is modified by addition to the cell of a nucleic acid of the invention; (b) culturing the modified cell to generate a plurality of modified cells; (c) measuring at least one metabolic parameter of the cell by monitoring the cell culture of step (b) in real time; and, (d) analyzing the data of step (c) to determine if the measured parameter differs from a comparable measurement in an unmodified cell under similar conditions, thereby identifying an engineered phenotype in the cell using real-time metabolic flux analysis. In one aspect, the genetic composition of the cell can be modified by a method comprising deletion of a sequence or modification of a sequence in the cell, or, knocking out the expression of a gene. In one aspect, the method can further comprise selecting a cell comprising a newly engineered phenotype. In another aspect, the method can comprise culturing the selected cell, thereby generating a new cell strain comprising a newly engineered phenotype.

The invention provides methods of increasing thermotolerance or thermostability of a polypeptide, enzyme, protein, e.g. structural or binding protein, polypeptide, the method comprising glycosylating a polypeptide, enzyme, protein, e.g. structural or binding protein, wherein the polypeptide, enzyme, protein, e.g. structural or binding protein comprises at least thirty contiguous amino acids of a polypeptide of the invention; or a polypeptide encoded by a nucleic acid sequence of the invention, thereby increasing thermotolerance or thermostability of the polypeptide, enzyme, protein, e.g. structural or

binding protein. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein specific activity can be thermostable or thermotolerant at a temperature in the range from greater than about 37°C to about 95°C.

The invention provides methods for overexpressing a recombinant polypeptide, enzyme, protein, e.g. structural or binding protein, in a cell comprising expressing a vector comprising a nucleic acid comprising a nucleic acid of the invention or a nucleic acid sequence of the invention, wherein the sequence identities are determined by analysis with a sequence comparison algorithm or by visual inspection, wherein overexpression is effected by use of a high activity promoter, a dicistronic vector or by gene amplification of the vector.

The invention provides methods of making a transgenic plant comprising the following steps: (a) introducing a heterologous nucleic acid sequence into the cell, wherein the heterologous nucleic sequence comprises a nucleic acid sequence of the invention, thereby producing a transformed plant cell; and (b) producing a transgenic plant from the transformed cell. In one aspect, the step (a) can further comprise introducing the heterologous nucleic acid sequence by electroporation or microinjection of plant cell protoplasts. In another aspect, the step (a) can further comprise introducing the heterologous nucleic acid sequence directly to plant tissue by DNA particle bombardment. Alternatively, the step (a) can further comprise introducing the heterologous nucleic acid sequence into the plant cell DNA using an *Agrobacterium tumefaciens* host. In one aspect, the plant cell can be a potato, corn, rice, wheat, tobacco, or barley cell.

The invention provides methods of expressing a heterologous nucleic acid sequence in a plant cell comprising the following steps: (a) transforming the plant cell with a heterologous nucleic acid sequence operably linked to a promoter, wherein the heterologous nucleic sequence comprises a nucleic acid of the invention; (b) growing the plant under conditions wherein the heterologous nucleic acids sequence is expressed in the plant cell. The invention provides methods of expressing a heterologous nucleic acid sequence in a plant cell comprising the following steps: (a) transforming the plant cell with a heterologous nucleic acid sequence operably linked to a promoter, wherein the heterologous nucleic sequence comprises a sequence of the invention; (b) growing the plant under conditions wherein the heterologous nucleic acids sequence is expressed in the plant cell.

The invention provides feeds or foods comprising a polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention. In one aspect, the invention provides a food, feed, a liquid, e.g., a beverage (such as a fruit juice or a beer), a bread or a dough or a bread product, or a beverage precursor (e.g., a wort), comprising a polypeptide of the invention. The invention provides food or nutritional supplements for an animal comprising a polypeptide of the invention, e.g., a polypeptide encoded by the nucleic acid of the invention.

In one aspect, the polypeptide in the food or nutritional supplement can be glycosylated. The invention provides edible enzyme delivery matrices comprising a polypeptide of the invention, e.g., a polypeptide encoded by the nucleic acid of the invention. In one aspect, the delivery matrix comprises a pellet. In one aspect, the polypeptide can be glycosylated. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein activity is thermotolerant. In another aspect, the polypeptide, enzyme, protein, e.g. structural or binding protein activity is thermostable.

The invention provides a food, a feed or a nutritional supplement comprising a polypeptide of the invention. The invention provides methods for utilizing a polypeptide, enzyme, protein, e.g. structural or binding protein, as a nutritional supplement in an animal diet, the method comprising: preparing a nutritional supplement containing a polypeptide, enzyme, protein, e.g. structural or binding protein, comprising at least thirty contiguous amino acids of a polypeptide of the invention; and administering the nutritional supplement to an animal. The animal can be a human, a ruminant or a monogastric animal. The polypeptide, enzyme, protein, e.g. structural or binding protein can be prepared by expression of a polynucleotide encoding the polypeptide, enzyme, protein, e.g. structural or binding protein in an organism selected from the group consisting of a bacterium, a yeast, a plant, an insect, a fungus and an animal. The organism can be selected from the group consisting of an *S. pombe*, *S. cerevisiae*, *Pichia pastoris*, *E. coli*, *Streptomyces* sp., *Bacillus* sp. and *Lactobacillus* sp.

The invention provides edible enzyme delivery matrix comprising thermostable recombinant polypeptide, enzyme, protein, e.g. structural or binding protein of the invention. The invention provides methods for delivering a polypeptide, enzyme, protein, e.g. structural or binding protein, supplement to an animal, the method comprising: preparing an edible enzyme delivery matrix in the form of pellets comprising a granulate edible carrier and thermostable recombinant polypeptide, enzyme, protein, e.g. structural or binding protein, wherein the pellets readily disperse the polypeptide, enzyme, protein,

e.g. structural or binding protein contained therein into aqueous media, and administering the edible enzyme delivery matrix to the animal. The recombinant polypeptide, enzyme, protein, e.g. structural or binding protein can comprise a polypeptide of the invention. The polypeptide, enzyme, protein, e.g. structural or binding protein can be glycosylated to provide thermostability at pelletizing conditions. The delivery matrix can be formed by pelletizing a mixture comprising a grain germ and a polypeptide, enzyme, protein, e.g. structural or binding protein. The pelletizing conditions can include application of steam. The pelletizing conditions can comprise application of a temperature in excess of about 80°C for about 5 minutes and the enzyme retains a specific activity of at least 350 to about 900 units per milligram of enzyme.

In one aspect, invention provides a pharmaceutical composition comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention, or a polypeptide encoded by a nucleic acid of the invention. In one aspect, the pharmaceutical composition acts as a digestive aid. In one aspect, the pharmaceutical of this invention is formulated as an edible delivery agent, or, a formulation of this invention comprises a tablet, a gel, a capsule or a geltab, or a lotion, a spray or a gel.

The invention provides methods for delivering an enzyme or protein supplement to an animal, the method comprising: (a) providing a cell that recombinantly generates a polypeptide of the invention, and (b) administering the cell or the recombinantly generated polypeptide to the animal. In one aspect, the cell is a plant cell, a bacterial cell, a yeast cell, an insect cell or an animal cell, or the cell is selected from the group consisting of a *Schizosaccharomyces* sp., *Saccharomyces* sp., *Pichia* sp., *Escherichia* sp., *Streptomyces* sp., *Bacillus* sp. and *Lactobacillus* sp., or the cell is *Saccharomyces pombe*, or the cell is *Saccharomyces cerevisiae*, or the cell is *Pichia pastoris*, or the cell is *Escherichia coli*, or the cell is *Bacillus cereus*.

The invention provides compositions comprising an encapsulated formulation, e.g., as a pharmaceutical, food, feed or dietary supplement encapsulated formulation, comprising at least one polypeptide of the invention.

The invention provides methods for making bioethanol, biomethanol, biopropanol and/or, bio-butanol comprising (A) contacting a composition comprising a cellooligsaccharide, an arabinoxylan oligomer, a lignin, a lignocellulose, a xylan, a glucan, a cellulose or a fermentable sugar with the polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention; (B) the method of (A), wherein the composition of comprises a plant, plant product or plant derivative, and optionally the plant

or plant product comprises cane sugar plants or plant products, beets or sugarbeets, wheat, corn, soybeans, potato, rice or barley, (C) the method of (A) or (B), wherein the polypeptide has activity comprising lignocellulosic enzyme and/or cellulase, endoglucanase, cellobiohydrolase, beta-glucosidase, xylanase, mannanse, β-xylosidase and/or arabinofuranosidase activity; or (D) the method of (A), (B) or (C), wherein the plant is a monocot or dicot, or the plant is a monocot corn, sugarcane, rice, wheat, barley, switchgrass or *Miscanthus*; or the plant is a dicot oilseed crop, soy, canola, rapeseed, flax, cotton, palm oil, sugar beet, peanut, tree, poplar or lupine; or (E) the method of (A), (B), (C) or (D), further comprising processing and/or formulating the bioethanol, biomethanol, biopropanol and/or, bio-butanol as a liquid fuel and/or a gas fuel, wherein optionally the fuel comprises a biofuel and/or a synthetic fuel, or the fuel comprises bioethanol, biomethanol, biopropanol and/or, bio-butanol; and/or a gasoline-ethanol, -methanol, -butanol and/or -propanol mix.

The invention provides methods for processing a biomass material comprising a cellooligsaccharide, an arabinoxylan oligomer, a lignin, a lignocellulose, a xylan, a glucan, a cellulose or a fermentable sugar comprising contacting the composition with the polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention, wherein optionally the biomass material comprising lignocellulose is derived from an agricultural crop, is a byproduct of a food or a feed production, is a lignocellulosic waste product, or is a plant residue or a waste paper or waste paper product, and optionally the polypeptide has activity comprising lignocellulosic enzyme and/or cellulase, endoglucanase, cellobiohydrolase, beta-glucosidase, xylanase, mannanse, β-xylosidase and/or arabinofuranosidase activity, and optionally the plant residue comprise stems, leaves, hulls, husks, corn or corn cobs, corn stover, hay, straw, wood, wood chips, wood pulp, wood waste and sawdust, and optionally the paper waste comprises discarded or used photocopy paper, computer printer paper, notebook paper, notepad paper, typewriter paper, newspapers, magazines, cardboard and paper-based packaging materials, and optionally the processing of the biomass material generates a biodiesel, or a bioalcohol, a bioethanol, biomethanol, biobutanol or biopropanol.

The details of one or more aspects of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, patents, patent applications, GenBank sequences and ATCC deposits, cited herein are hereby expressly incorporated by reference for all purposes.

## BRIEF DESCRIPTION OF DRAWINGS

The following drawings are illustrative of aspects of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is a block diagram of a computer system.

Figure 2 is a flow diagram illustrating one aspect of a process for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 3 is a flow diagram illustrating one aspect of a process in a computer for determining whether two sequences are homologous.

Figure 4 is a flow diagram illustrating one aspect of an identifier process 300 for detecting the presence of a feature in a sequence.

Like reference symbols in the various drawings indicate like elements.

## DETAILED DESCRIPTION

The invention provides isolated, synthetic and recombinant polypeptides, including enzymes, structural proteins and binding proteins, polynucleotides encoding these polypeptides, and methods of making and using these polynucleotides and polypeptides. The polypeptides of the invention, and the polynucleotides encoding the polypeptides of the invention, encompass many classes of enzymes, structural proteins and binding proteins. In one aspect, the enzymes and proteins of the invention comprise, e.g. aldolases, alpha-galactosidases, amidases, e.g. secondary amidases, amylases, catalases, carotenoid pathway enzymes, dehalogenases, endoglucanases, epoxide hydrolases, esterases, hydrolases, glucosidases, glycosidases, inteins, isomerases, laccases, lipases, monooxygenases, nitroreductases, nitrilases, P450 enzymes, pectate lyases, phosphatases, phospholipases, phytases, polymerases, proteases, peptidases and xylanases, which are more specifically described below. The invention also provides isolated, synthetic and recombinant polypeptides, including enzymes, structural proteins and binding proteins, polynucleotides encoding these polypeptides, having the activities described in Table 1, Table 3 and/or Table 4, and in another aspect, with or without signal sequences as described in Table 2.

### Aldolases

In one aspect, the invention provides aldolases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an aldolase activity,

including thermostable and thermotolerant aldolase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides. In one aspect, the aldolase activity comprises catalysis of the formation of a carbon-carbon bond. In one aspect, the aldolase activity comprises an aldol condensation. The aldol condensation can have an aldol donor substrate comprising an acetaldehyde and an aldol acceptor substrate comprising an aldehyde. The aldol condensation can yield a product of a single chirality. In one aspect, the aldolase activity is enantioselective. The aldolase activity can comprise a 2-deoxyribose-5-phosphate aldolase (DERA) activity. The aldolase activity can comprise catalysis of the condensation of acetaldehyde as donor and a 2(R)-hydroxy-3-(hydroxy or mercapto)-propionaldehyde derivative to form a 2-deoxysugar. The aldolase activity can comprise catalysis of the condensation of acetaldehyde as donor and a 2-substituted acetaldehyde acceptor to form a 2,4,6-trideoxyhexose via a 4-substituted-3-hydroxybutanal intermediate. The aldolase activity can comprise catalysis of the generation of chiral aldehydes using two acetaldehydes as substrates. The aldolase activity can comprises enantioselective assembling of chiral β,δ-dihydroxyheptanoic acid side chains. The aldolase activity can comprise enantioselective assembling of the core of [R-(R*,R*)]-2-(4-fluorophenyl)-b,d-dihydroxy-5-(1-methylethyl)-3-phenyl-4-(phenylamino)-carbonyl]-1H-pyrrole-1-heptanoic acid (Atorvastatin, or LIPITOR™), rosuvastatin (CRESTOR™) and/or fluvastatin (LESCOL™). The aldolase activity can comprise, with an oxidation step, synthesis of a 3R,5S-6-chloro-2,4,6-trideoxy-erythro-hexonolactone.

Alpha-galactosidases

In one aspect, the invention provides alpha-galactosidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an alpha-galactosidase activity, including thermostable and thermotolerant alpha-galactosidase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

An alpha galactosidase hydrolyses the non-reducing terminal alpha 1-3,4,6 linked galactose from poly- and oligosaccharides. These saccharides are commonly found in legumes and are difficult to digest. As such, alpha-galactosidases can be used as a digestive aid to break down raffinose, stachyose, and verbascose, found in such foods as beans and other gassy foods.

Amidases

In one aspect, the invention provides amidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an amidase activity, including thermostable and thermotolerant amidase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides. In one aspect, the amidases of the invention are used in the removal of arginine, phenylalanine or methionine from the N-terminal end of peptides in peptide or peptidomimetic synthesis. In one aspect, the enzyme of the invention, e.g. an amidase, is selective for the L, or "natural" enantiomer of the amino acid derivatives and is therefore useful for the production of optically active compounds. These reactions can be performed in the presence of the chemically more reactive ester functionality, a step which is very difficult to achieve with nonenzymatic methods. The enzyme is also able to tolerate high temperatures (at least 70° C), and high concentrations of organic solvents (>40% DMSO), both of which cause a disruption of secondary structure in peptides, which enables cleavage of otherwise resistant bonds.

Secondary amidases

In one aspect, the invention provides secondary amidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a secondary amidase activity, including thermostable and thermotolerant secondary amidase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Secondary amidases include a variety of useful enzymes including peptidases, proteases, and hydantoinases. This class of enzymes can be used in a range of commercial applications. For example, secondary amidases can be used to: 1) increase flavor in food, in particular cheese (known as enzyme ripened cheese); 2) promote bacterial and fungal killing; 3) modify and de-protect fine chemical intermediates 4) synthesize peptide bonds; 5) and carry out chiral resolutions. Particularly, there is a need in the art for an enzyme capable of hydrolyzing Cephalosporin C.

Amylases

In one aspect, the invention provides amylases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect,

the invention is directed to polypeptides, e.g., enzymes, having an amylase activity, including thermostable and thermotolerant amylase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In one aspect, the polypeptides of the invention can be used as amylases, for example, alpha amylases or glucoamylases, to catalyze the hydrolysis of starch into sugars. In one aspect, the invention is directed to polypeptides having thermostable amylase activity, such as alpha amylases or glucoamylase activity, e.g., a 1,4-alpha-D-glucan glucohydrolase activity. In one aspect, the polypeptides of the invention can be used as amylases, for example, alpha amylases or glucoamylases, to catalyze the hydrolysis of starch into sugars, such as glucose. The invention is also directed to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences of the invention as well as recombinant methods for producing the polypeptides of the invention. The invention is also directed to the use of amylases of the invention in starch conversion processes, including production of high fructose corn syrup (HFCS), bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, dextrose, and dextrose syrups and the like.

Commercially, glucoamylases are used to further hydrolyze cornstarch, which has already been partially hydrolyzed with an alpha-amylase. The glucose produced in this reaction may then be converted to a mixture of glucose and fructose by a glucose isomerase enzyme. This mixture, or one enriched with fructose, is the high fructose corn syrup commercialized throughout the world. In general, starch to fructose processing consists of four steps: liquefaction of granular starch, saccharification of the liquefied starch into dextrose, purification, and isomerization to fructose. The object of a starch liquefaction process is to convert a concentrated suspension of starch polymer granules into a solution of soluble shorter chain length dextrins of low viscosity.

The amylases of the invention can be used in automatic dish wash (ADW) products and laundry detergent. In ADW products, the amylase will function at pH 10-11 and at 45-60°C in the presence of calcium chelators and oxidative conditions. For laundry, activity at pH 9-10 and 40°C in the appropriate detergent matrix will be required. Amylases are also useful in textile desizing, brewing processes, starch modification in the paper and pulp industry and other processes described in the art.

Amylases can be used commercially in the initial stages (liquefaction) of starch processing; in wet corn milling; in alcohol production; as cleaning agents in detergent matrices; in the textile industry for starch desizing; in baking applications; in the beverage

industry; in oilfields in drilling processes; in inking of recycled paper and in animal feed. Amylases are also useful in textile desizing, brewing processes, starch modification in the paper and pulp industry and other processes.

Carotenoid pathway enzymes

The invention provides novel enzymes, and the polynucleotides encoding them, involved in carotenoid (such as lycopenes and luteins), astaxanthin and/or isoprenoid synthesis. The invention also provides novel genes in the carotenoid, astaxanthin and isoprenoid biosynthetic pathways comprising at least one enzyme of the invention. For example, alternative aspects, the invention provides one or more nucleic acid coding sequences (CDSs, or ORFs) encoding all, or at least one, enzyme(s) involved in a desired biosynthetic pathway for carotenoids, astaxanthins and/or isoprenoids. The nucleic acid coding sequence(s) can be expressed through an expression plasmid, vector, engineered virus or any episomal expression system, or, can be integrated into the genome of the host cell. In one aspect, the enzyme(s) involved in the biosynthetic pathway system comprise a novel combination of enzymes. In another aspect, the enzyme(s) involved in the biosynthetic pathway system comprise at least one novel enzyme of the invention – where nucleic acids used in the system encode a novel enzyme of the invention.

Carotenoids are natural pigments which have antioxidant and anti-carcinogenic activity. They are free radical scavengers, and as such, strong antioxidants. Carotenoids have a conjugated backbone structure and are very rigid molecules, having a backbone consisting of 9 to 11 alternating single/double bonds and have very similar electro-optical properties as polyacetylene. Astaxanthins are abundant naturally occurring carotenoids. They contain an internal unit similar to beta-carotene but have two terminal carbonyl and hydroxyl functionalities. These compounds are useful for food and feed supplements, colorants, neutraceuticals, cosmetic and pharmaceutical needs. Isoprenoids are compounds biosynthesized from or containing isoprene (unsaturated branched chain five-carbon hydrocarbon) units, including terpenes, carotenoids, fat soluble vitamins, ubiquinone, rubber, and some steroids. Biosynthetic pathways for carotenoids, astaxanthins and isoprenoids are known; most of these published pathways are derived from one organism or a combination of genes from a few species.

Catalases

In one aspect, the invention provides catalases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect,

the invention is directed to polypeptides, e.g., enzymes, having a catalase activity, including thermostable and thermotolerant catalase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In processes where hydrogen peroxide is a by-product, catalases of the invention can be used to destroy or detect hydrogen peroxide, e.g., in production of glyoxylic acid and in glucose sensors. Also, in processes where hydrogen peroxide is used as a bleaching or antibacterial agent, catalases of the invention can be used to destroy residual hydrogen peroxide, e.g. in contact lens cleaning, in bleaching steps in pulp and paper production, and in the pasteurization of dairy products. Further, such catalases of the invention can be used as catalysts for oxidation reactions, e.g. epoxidation and hydroxylation.

Dehalogenases

In one aspect, the invention provides dehalogenases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a dehalogenase activity, including thermostable and thermotolerant dehalogenase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Environmental pollutants consist of a large quantity and variety of chemicals; many of these are toxic, environmental hazards that were designated in 1979 as priority pollutants by the U.S. Environmental Protection Agency. Microbial and enzymatic biodegradation is one method for the elimination of these pollutants. Accordingly, methods have been designed to treat commercial wastes and to bioremediate polluted environments via microbial and related enzymatic processes. Unfortunately, many chemical pollutants are either resistant to microbial degradation or are toxic to potential microbial-degraders when present in high concentrations and certain combinations.

Dehalogenases, e.g. haloalkane dehalogenases, of the invention can cleave carbon-halogen bonds in haloalkanes and halocarboxylic acids by hydrolysis, thus converting them to their corresponding alcohols. This reaction can be used for detoxification involving haloalkanes, such as ethylchloride, methylchloride, and 1,2-dichloroethane (e.g., detoxification of toxic composition, e.g., pesticides, poisons, chemical warfare agents and the like comprising haloalkanes).

The present invention provides a number of dehalogenase enzymes useful in bioremediation having improved enzymatic characteristics. The polynucleotides and

polynucleotide products of the invention are useful in, for example, groundwater treatment involving transformed host cells containing a polynucleotide or polypeptide of the invention (e.g., the bacteria Xanthobacter autotrophicus) and the haloalkane 1,2-dichlorethane as well as removal of polychlorinated biphenyls (PCB's) from soil sediment.

The haloalkane dehalogenase of the invention are useful in carbon-halide reduction efforts. The enzymes of the invention initiate the degradation of haloalkanes. Alternatively, host cells containing a dehalogenase polynucleotide or polypeptide of the invention can feed on the haloalkanes and produce the detoxifying enzyme.

Endoglucanases

In one aspect, the invention provides endoglucanases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an endoglucanase activity, including thermostable and thermotolerant endoglucanase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In one aspect, the enzymes of the invention have a glucanase, e.g., an endoglucanase, activity, e.g., catalyzing hydrolysis of internal endo- β-1,4- and/or β-1,3-glucanase linkages. In one aspect, the endoglucanase activity (e.g., endo-1,4-beta-D-glucan 4-glucano hydrolase activity) comprises hydrolysis of 1,4- and/or β-1,3- beta-D-glycosidic linkages in cellulose, cellulose derivatives (e.g., carboxy methyl cellulose and hydroxy ethyl cellulose) lichenin, beta-1,4 bonds in mixed beta-1,3 glucans, such as cereal beta-D-glucans or xyloglucans and other plant material containing cellulosic parts.

Endoglucanases of the invention (e.g., endo-beta-1,4-glucanases, EC 3.2.1.4; endo-beta-1,3(1)-glucanases, EC 3.2.1.6; endo-beta-1,3-glucanases, EC 3.2.1.39) can hydrolyze internal β-1,4- and/or β-1,3- glucosidic linkages in cellulose and glucan to produce smaller molecular weight glucose and glucose oligomers. Glucans are polysaccharides formed from 1,4-β- and/or 1,3-glycoside-linked D-glucopyranose. Endoglucanases of the invention can be used in the food industry, for baking and fruit and vegetable processing, breakdown of agricultural waste, in the manufacture of animal feed, in pulp and paper production, textile manufacture and household and industrial cleaning agents. Endoglucanases are produced by fungi and bacteria.

Beta-glucans are major non-starch polysaccharides of cereals. The glucan content can vary significantly depending on variety and growth conditions. The physicochemical

properties of this polysaccharide are such that it gives rise to viscous solutions or even gels under oxidative conditions. In addition glucans have high water-binding capacity. All of these characteristics present problems for several industries including brewing, baking, animal nutrition. In brewing applications, the presence of glucan results in wort filterability and haze formation issues. In baking applications (especially for cookies and crackers), glucans can create sticky doughs that are difficult to machine and reduce biscuit size. In addition, this carbohydrate is implicated in rapid rehydration of the baked product resulting in loss of crispiness and reduced shelf-life. For monogastric animal feed applications with cereal diets, beta-glucan is a contributing factor to viscosity of gut contents and thereby adversely affects the digestibility of the feed and animal growth rate. For ruminant animals, these beta-glucans represent substantial components of fiber intake and more complete digestion of glucans would facilitate higher feed conversion efficiencies. It is desirable for animal feed endoglucanases to be active in the animal stomach.

Endoglucanases of the invention can be used in the digestion of cellulose, a beta-1,4-linked glucan found in all plant material. Cellulose is the most abundant polysaccharide in nature. Enzymes of the invention that digest cellulose have utility in the pulp and paper industry, in textile manufacture and in household and industrial cleaning agents.

The following assays can be used to demonstrate/confirm the enzymatic activity of exemplary enzymes of the invention or to determine if a polypeptide has the requisite enzyme (e.g., glucanase, mannanase, or xylanase or e.g., an alkaline endoglucanase/cellulase) activity to be within the scope of the invention.

Cellulase activity assay: BCA reducing ends assay

This assay measures the amount of reducing ends produced during the enzymatic degradation of carboxymethylcellulose (CMC) in a high throughput multiple sample 96-well format.

Materials:

Substrate solutions:

1% CMC - Dissolve 1 gm CMC in 100 ml 50 mM Britton-Robinson buffer at pH ~4, heat CMC solution in boiling water bath, while mixing, for 20-40 minutes until it dissolves (solution will still appear slightly milky, but translucent). Adjust to desired pH with 1M NaOH or HCl.

Solution A:

64 mg/ml sodium carbonate monohydrate

24 mg/ml sodium bicarbonate

1.95 mg/ml BCA (4,4'-dicarboxy-2,2'- biquinoline disodium salt (Sigma Chemical cat # D-8284)

Add above to dH2O,

Might need to dissolve the BCA by heating, don't heat more than ~80 C.

Solution B:

1.24 mg/ml cupric sulfate pentahydrate

1.26 mg/ml L-serine

Add above to dH2O

Working reagent:

1:1 of solutions A & B, make fresh working reagent mixture every day (usually only make enough for each assay), make fresh Solutions A & B every week.

Glucose stock solution:

10 mM Glucose in dH2O. 0.2 um filter, store at 4C.

Glucose standards:

Dilute the 10 mM Glucose stock in 1% CMC at desired pH; to a final concentration of 0, 100, 200, 300, 400, 500 uM. Since the curve is determined by adding 10 ul of the standards to the working reagent it works out to 0-0.005 umole glucose/well. The standard curve needs to be generated for each plate of sample time-points, as the heating cycle can affect the amount of signal observed.

Methods:

Set-Up:

Aliquot 1 ml of substrate solution (1% CMC) into deep-well plate (if using ambient Temp) or Acme-tubes in hot-block, equilibrate to desired temperature (~5 min) in heat block or heated water bath.

While solution is equilibrating, make 10 ml of the working reagent and aliquot 100 ul into 96 well PCR-plate. Set plate on ice.

Reaction/Sampling:

After temperature equilibration is complete, add enzyme solution to substrate solution. Mix immediately by pipetting up/down. Immediately aliquot 10-ul into PCR-plate (this is t=0, zero time point). Aliquot 10-ul into PCR-plate at each desired time point (e.g. 0, 2, 5, 10, 15, 20, 30 minutes).

Save the last row on the plate for addition of 10ul of glucose standards (I.e. wells should only have the 100-ul working reagent in them)

Assay Color Development:

When all time points are collected and standards are added, cover plate and heat to 100 C for 10 min using PCR machine. Cool plate on ice for 5-10 min (or set PCR machine to 1C for 10 min).

Add 100 ul H2O to wells. Mix. Aliquot 100 ul of mixture into clear flat bottomed 96-well plate and read absorbance at 560 nm.

Generate standard curve:

Plot the A560 vs. umole glucose from the wells containing the glucose standards. Use linear regression to calculate the slope ($S_{std}$).

Generate graph of reaction slope:

Plot A560 vs. time-points. Zero each sample's time points against its own T=0 (i.e. subtract sample's T=0 absorbance value from all other time-points of same sample).

Generate the slope ($S_{rxn}$) for each set of sample time-points (A560/time).

Activity Determination:

Divide $S_{rxn}$ by the $S_{std}$, and multiply by 100 (as the umole product detected is the amount of reducing ends in the 10-ul used in the assay, not the total amount generated in the 1ml enzyme reaction).

Specific Activity Determination:

Divide the Activity (in units of umole/min) by the total mg of protein added in the 1-ml reaction. Determine the protein concentration by Bradford or similar assay.

Divide the protein concentration by any dilutions used.

Multiply by the volume (in ml) used in the reaction.

All points should be done in duplicate with triplicate being better.

In another aspect, enzymes are tested to determine if they are within the scope of the invention using a glucanase substrate (many of which are well known in the art), e.g., barley and/or oat glucan, and/or galactomannan-comprising compositions.

Specific activity of the enzymes of the invention can be determined on 1% substrate in 50 mM sodium acetate buffer pH 5.3, at 37°C using the BCA reducing sugar assay (see above). 1 unit (U) of glucanase activity = 1 $\mu mol/min^{-1}$ glucose reducing equivalents released at 37°C, pH 5.3.

- Barley Beta Glucan (BBG): 30 U/mg
- Oat Beta Glucan (OBG): 38 U/mg
- Carboxymethylcellulose (CMC): 40 U/mg
- Carob Galactomannan: 0.3 U/mg

Epoxide hydrolases

In one aspect, the invention provides epoxide hydrolases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an epoxide hydrolase activity, including thermostable and thermotolerant epoxide hydrolase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides. The polypeptides of the invention can be used as epoxide hydrolases to catalyze the hydrolysis of epoxides and arene oxides to their corresponding diols.

Epoxide hydrolases catalyze the hydrolysis of epoxides and arene oxides to their corresponding diols. Epoxide hydrolases from microbial sources are highly versatile biocatalysts for the asymmetric hydrolysis of epoxides on a preparative scale. Besides kinetic resolution, which furnishes the corresponding vicinal diol and remaining non-hydrolyzed epoxide in nonracemic form, enantioconvergent processes are possible. These are highly attractive as they lead to the formation of a single enantiomeric diol from a racemic oxirane.

Microsomal epoxide hydrolases are biotransformation enzymes that catalyze the conversion of a broad array of xenobiotic epoxide substrates to more polar diol metabolites, see, e.g., Omiecinski (2000) Toxicol. Lett. 112-113:365-370. Microsomal epoxide hydrolases catalyze the addition of water to epoxides in a two-step reaction involving initial attack of an active site carboxylate on the oxirane to give an ester intermediate followed by hydrolysis of the ester. Soluble epoxide hydrolase play a role in the biosynthesis of inflammation mediators.

Epoxide hydrolases of the invention can be used in the detoxification of epoxides or in the biosynthesis of hormones. Additionally, epoxide hydrolases of the invention can efficiently process several substrates, leading to enantiomerically enriched–epoxides (the unreacted enantiomer) and/or to the corresponding vicinal diols.

Esterases

In one aspect, the invention provides esterases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an esterase activity, including thermostable and thermotolerant esterase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Many esterases are known and have been discovered in a broad variety of organisms, including bacteria, yeast and higher animals and plants. A principal example of esterases are the lipases, which are used in the hydrolysis of lipids, acidolysis (replacement of an esterified fatty acid with a free fatty acid) reactions, transesterification (exchange of fatty acids between triglycerides) reactions, and in ester synthesis. The major industrial applications for lipases include: the detergent industry, where they are employed to decompose fatty materials in laundry stains into easily removable hydrophilic substances; the food and beverage industry where they are used in the manufacture of cheese, the ripening and flavoring of cheese, as antistaling agents for bakery products, and in the production of margarine and other spreads with natural butter flavors; in waste systems; and in the pharmaceutical industry where they are used as digestive aids.

Alternatively, esterases of the invention can be used in detergent compositions. In one aspect, the esterase can be a nonsurface-active esterase. In another aspect, the esterase can be a surface-active esterase. The esterase can be formulated in a non-aqueous liquid composition, a cast solid, a granular form, a particulate form, a compressed tablet, a gel form, a paste or a slurry form.

In another aspect, the invention provides fabrics or clothing comprising an esterase of the invention. In another aspect, esterases of the invention are used to treat a lipid-containing fabric.

In another aspect, the invention provides foods and drinks comprising an esterase of the invention. The invention also provides cheeses comprising an esterase of the invention. Additionally, the invention provides methods for the manufacture of cheese comprising the following steps: (a) providing a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention; (b) providing a cheese precursor; and (c) contacting the polypeptide of step (a) with the precursor of step (b) under condition wherein the esterase can catalyze cheese manufacturing processes. In one aspect, the method can comprise the process of ripening and flavoring of cheese.

In another aspect, the invention provides margarines and spreads comprising an enzyme of the invention. The invention provides methods for production of margarine or other spreads with natural butter flavors comprising the following steps: (a) providing a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention; (b) providing a margarine or a spread precursor; and (c) contacting the polypeptide of step (a) with the precursor of step (b) under condition wherein the esterase can catalyze processes involved in margarine or spread production.

The invention provides methods for treating solid or liquid waste products comprising the following steps: (a) providing a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention; (b) providing a solid or a liquid waste; and (c) contacting the polypeptide of step (a) and the waste of step (b) under conditions wherein the polypeptide can treat the waste. The invention provides solid or liquid waste products comprising a polypeptide of the invention.

The invention provides methods for aiding digestion in a mammal comprising (a) providing a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention; (b) providing a composition comprising a substrate for the polypeptide of step (a); (c) feeding or administering to the mammal the polypeptide of step (a) with a feed or food comprising a substrate for the polypeptide of step (a), thereby helping digestion in the mammal. In one aspect, the mammal is a human.

The invention provides pharmaceutical compositions comprising a polypeptide and/or a nucleic acid of the invention, e.g., a pharmaceutical composition for use as a digestive aid in a mammal comprising a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention. In one aspect, the mammal comprises a human. The enzymes of the invention are used in the manufacture of medicaments.

The invention provides bakery products comprising a polypeptide of the invention. The invention provides antistaling agents for bakery products comprising a polypeptide having an esterase activity, wherein the polypeptide comprises a polypeptide of the invention, or, a polypeptide encoded by a nucleic acid of the invention.

The invention provides methods for hydrolyzing, breaking up or disrupting a ester-comprising composition comprising the following steps: (a) providing a polypeptide

of the invention having an esterase activity, or a polypeptide encoded by a nucleic acid of the invention; (b) providing a composition comprising a protein; and (c) contacting the polypeptide of step (a) with the composition of step (b) under conditions wherein the esterase hydrolyzes, breaks up or disrupts the ester-comprising composition.

Alternatively, the invention provides methods for liquefying or removing ester-comprising compositions comprising the following steps: (a) providing a polypeptide of the invention having an esterase activity, or a polypeptide encoded by a nucleic acid of the invention; (b) providing a composition comprising a protein; and (c) contacting the polypeptide of step (a) with the composition of step (b) under conditions wherein esterase removes or liquefies the ester-comprising compositions.

Hydrolases

In one aspect, the invention provides hydrolases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a hydrolase activity, e.g., an esterase, acylase, lipase, phospholipase or protease (including peptidase) activity, including thermostable and thermotolerant hydrolase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides. The hydrolase activities of the polypeptides and peptides of the invention include esterase activity, lipase activity (hydrolysis of lipids), acidolysis reactions (to replace an esterified fatty acid with a free fatty acid), transesterification reactions (exchange of fatty acids between triglycerides), ester synthesis, ester interchange reactions, phospholipase activity (e.g., phospholipase A, B, C and D activity, patatin activity, lipid acyl hydrolase (LAH) activity) and protease activity (hydrolysis of peptide bonds). The polypeptides of the invention can be used in a variety of pharmaceutical, agricultural and industrial contexts, including the manufacture of cosmetics and nutraceuticals.

In one aspect, the polypeptides of the invention are used in the biocatalytic synthesis of structured lipids (lipids that contain a defined set of fatty acids distributed in a defined manner on the glycerol backbone), including cocoa butter alternatives (CBA), lipids containing poly-unsaturated fatty acids (PUFAs), diacylglycerides, e.g., 1,3-diacyl glycerides (DAGs), monoglycerides, e.g., 2-monoglycerides (MAGs) and triacylglycerides (TAGs). In one aspect, the polypeptides of the invention are used to modify oils, such as fish, animal and vegetable oils, and lipids, such as poly-unsaturated fatty acids. The hydrolases of the invention having lipase activity can modify oils by hydrolysis, alcoholysis, esterification, transesterification and/or interesterification. The

methods of the invention can use lipases with defined regio-specificity or defined chemoselectivity in biocatalytic synthetic reactions. In another aspect, the polypeptides of the invention are used to synthesize enantiomerically pure chiral products.

Additionally, the polypeptides of the invention can be used in food processing, brewing, bath additives, alcohol production, peptide synthesis, enantioselectivity, hide preparation in the leather industry, waste management and animal degradation, silver recovery in the photographic industry, medical treatment, silk degumming, biofilm degradation, biomass conversion to a biofuel such as bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel and the like, biodefense, antimicrobial agents and disinfectants, personal care and cosmetics, biotech reagents, in increasing starch yield from corn wet milling and pharmaceuticals such as digestive aids and anti-inflammatory (anti-phlogistic) agents.

The major industrial applications for hydrolases, e.g., esterases, lipases, phospholipases and proteases, include the detergent industry, where they are employed to decompose fatty materials in laundry stains into easily removable hydrophilic substances; the food and beverage industry where they are used in the manufacture of cheese, the ripening and flavoring of cheese, as antistaling agents for bakery products, and in the production of margarine and other spreads with natural butter flavors; in waste systems; and in the pharmaceutical industry where they are used as digestive aids.

Oils and fats an important renewable raw material for the chemical industry. They are available in large quantities from the processing of oilseeds from plants like rice bran oil, rapeseed (canola), sunflower, olive, palm or soy. Other sources of valuable oils and fats include fish, restaurant waste, and rendered animal fats. These fats and oils are a mixture of triglycerides or lipids, i.e. fatty acids (FAs) esterified on a glycerol scaffold. Each oil or fat contains a wide variety of different lipid structures, defined by the FA content and their regiochemical distribution on the glycerol backbone. These properties of the individual lipids determine the physical properties of the pure triglyceride. Hence, the triglyceride content of a fat or oil to a large extent determines the physical, chemical and biological properties of the oil. The value of lipids increases greatly as a function of their purity. High purity can be achieved by fractional chromatography or distillation, separating the desired triglyceride from the mixed background of the fat or oil source. However, this is costly and yields are often limited by the low levels at which the triglyceride occurs naturally. In addition, the purity of the product is often compromised

by the presence of many structurally and physically or chemically similar triglycerides in the oil.

An alternative to purifying triglycerides or other lipids from a natural source is to synthesize the lipids. The products of such processes are called structured lipids because they contain a defined set of fatty acids distributed in a defined manner on the glycerol backbone. The value of lipids also increases greatly by controlling the fatty acid content and distribution within the lipid. Lipases can be used to affect such control.

Phospholipases are enzymes that hydrolyze the ester bonds of phospholipids. Corresponding to their importance in the metabolism of phospholipids, these enzymes are widespread among prokaryotes and eukaryotes. The phospholipases affect the metabolism, construction and reorganization of biological membranes and are involved in signal cascades. Several types of phospholipases are known which differ in their specificity according to the position of the bond attacked in the phospholipid molecule. Phospholipase A1 (PLA1) removes the 1-position fatty acid to produce free fatty acid and 1-lyso-2-acylphospholipid. Phospholipase A2 (PLA2) removes the 2-position fatty acid to produce free fatty acid and 1-acyl-2-lysophospholipid. PLA1 and PLA2 enzymes can be intra- or extra-cellular, membrane-bound or soluble. Intracellular PLA2 is found in almost every mammalian cell. Phospholipase C (PLC) removes the phosphate moiety to produce 1,2 diacylglycerol and phospho base. Phospholipase D (PLD) produces 1,2-diacylglycerophosphate and base group. PLC and PLD are important in cell function and signaling. Patatins are another type of phospholipase thought to work as a PLA.

In general, enzymes, including hydrolases such as esterases, lipases and proteases, are active over a narrow range of environmental conditions (temperature, pH, etc.), and many are highly specific for particular substrates. The narrow range of activity for a given enzyme limits its applicability and creates a need for a selection of enzymes that (a) have similar activities but are active under different conditions or (b) have different substrates. For instance, an enzyme capable of catalyzing a reaction at 50°C may be so inefficient at 35°C, that its use at the lower temperature will not be feasible. For this reason, laundry detergents generally contain a selection of proteolytic enzymes (e.g., polypeptides of the invention), allowing the detergent to be used over a broad range of wash temperature and pH. In view of the specificity of enzymes and the growing use of hydrolases in industry, research, and medicine, there is an ongoing need in the art for new enzymes and new enzyme inhibitors.

Proteases

In one aspect, the invention provides proteases and peptidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. Enzymes of this invention can have any protease activity, including a peptidase and/or a proteinase activity; for example, an exemplary protease activity of the invention can comprise catalysis of the hydrolysis of peptide bonds. The proteases of the invention can catalyze peptide hydrolysis reactions in both directions. The direction of the reaction can be determined, e.g., by manipulating substrate and/or product concentrations, temperature, selection of protease and the like. The protease activity can comprise an endoprotease activity and/or an exoprotease activity. The protease activity can comprise a protease activity, e.g., a carboxypeptidase activity, a dipeptidylpeptidase or an aminopeptidase activity, a serine protease activity, a metalloproteinase activity, a cysteine protease activity and/or an aspartic protease activity. In one aspect, protease activity can comprise activity the same or similar to a chymotrypsin, a trypsin, an elastase, a kallikrein and/or a subtilisin activity.

Assays that can be used to determine if a polypeptide has a protease activity can comprise use of a variety of pNA (para-nitroanalide) linked small peptide substrates as well as protein substrates, such as casein, gelatin, corn zein, soybean trypsin inhibitor, soybean lectin, and wheat germ lectin. For the small peptide substrate assays, hydrolysis of the terminal peptide bond liberates the pNA group and causes an increase in absorbance at 410nm. To monitor activity on the protein substrates, incubation of the protease and substrate at 37°C can be followed by monitoring the increase in fluorescence from an intramolecularly quenched substrate, by O-pthaldialdehyde (OPA) analysis, where in the presence of BME, OPA reacts with free amino ends to produce a fluorescent imidazole that can be detected using a standard fluorescence plate reader, or by SDS-PAGE analysis, where protease activity is indicated by the reduction or disappearance of substrate band(s).

Proteinase activity can also be determined on casein, gelatin, or corn zein using zymograms: zymogram gels contain the enzyme substrate (e.g., alpha-zein) embedded within the gel matrix. If a protease has activity on the zein in the gel, a clearing zone will be produced within an otherwise blue background following electrophoresis, renaturation, development, and staining steps. The clearing zone corresponds to the position of the protease in the gel.

The polypeptides of the invention include proteases in an active or inactive form. For example, the polypeptides of the invention include proproteins before "maturation" or

processing of prepro sequences, e.g., by a proprotein-processing enzyme, such as a proprotein convertase to generate an "active" mature protein. The polypeptides of the invention include proteases inactive for other reasons, e.g., before "activation" by a post-translational processing event, e.g., an endo- or exo-peptidase or proteinase action, a phosphorylation event, an amidation, a glycosylation or a sulfation, a dimerization event, and the like.

The polypeptides of the invention include all active forms, including active subsequences, e.g., catalytic domains or active sites, e.g., of a protease (or any other enzyme) of this invention. In one aspect, the invention provides catalytic domains or active sites. In one aspect, the invention provides a peptide or polypeptide comprising or consisting of an active site domain; domains can be predicted through use of the database Pfam, which is a large collection of multiple sequence alignments and hidden Markov models covering many common protein families, The Pfam protein families database, A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, and E.L.L. Sonnhammer, Nucleic Acids Research, 30(1):276-280, 2002); Finn, et al., (2006) Nucleic Acids Res., Database Issue 34:D247-D251. Profile hidden Markov models (profile HMMs) can be used to do database (sequence) searching using statistical descriptions of a sequence family's consensus; HMMER uses profile HMMs, and can be useful in situations like working with an evolutionarily diverse protein family (a BLAST search with any individual sequence may not find the rest of the sequences in the family).

Methods for identifying "prepro" domain sequences and signal sequences are well known in the art, see, e.g., Van de Ven (1993) Crit. Rev. Oncog. 4(2):115-136. For example, to identify a prepro sequence, the protein is purified from the extracellular space and the N-terminal protein sequence is determined and compared to the unprocessed form.

Glucosidases

In one aspect, the invention provides glucosidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a glucosidase activity, including thermostable and thermotolerant glucosidase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Alpha-glucosidases of the invention can catalyze the hydrolysis of starches into sugars. Alpha-glucosidases can hydrolyze terminal non-reducing 1,4 or 1,6 linked α-D-glucose residues in starch, with release of α-D-glucose.

Alpha-glucosidases of the invention can be used commercially in the stages liquefaction and saccharification of starch processing; in wet corn milling; in alcohol production; as cleaning agents in detergent matrices; in the textile industry for starch desizing; in baking applications; in the beverage industry; in oilfields in drilling processes; in inking of recycled paper and in animal feed. Alpha-glucosidases of the invention are also useful in textile desizing, brewing processes, starch modification in the paper and pulp industry and other processes.

Glycosidases

In one aspect, the invention provides glycosidases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a glycosidase activity, including thermostable and thermotolerant glycosidase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides. Glycosidase enzymes of the invention can have more specific activity as glucosidases, α-galactosidases, β-galactosidases, β-mannosidases, β-mannanases, endoglucanases, and pullulanases.

α-galactosidases of the invention can catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose. β-mannanases of the invention can catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising mannose groups. β-mannosidases of the invention can hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccaharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of β-1, 4 linked mannose backbone with a-1, 6 linked galactose sidechains. The enzymes required for the degradation of guar are β-mannanase, β-mannosidase and α-galactosidase. β-mannanase hydrolyses the mannose backbone internally and β-mannosidase hydrolyses non-reducing, terminal mannose residues. α-galactosidase hydrolyses α-linked galactose groups.

Galactomannan polysaccharides and the enzymes of the invention that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a need exists for thermostable galactomannases that are active in extreme conditions associated with oil drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is merit to eliminating raffinose from raw beet sugar. α-Galactosidase has also been used as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

β-Galactosidases of the invention can be used for the production of lactose-free dietary milk products. Additionally, β-galactosidases of the invention can be used for the enzymatic synthesis of oligosaccharides via transglycosylation reactions.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme of the invention can hydrolyze α-1, 6-glucosidic linkages on these polymers. Starch degradation for the production or sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with α-amylase, and the second stage, or saccharification stage, is performed by β-amylase with pullalanase added as a debranching enzyme, to obtain better yields.

Endoglucanases of the invention can be used in a variety of industrial applications. For instance, the endoglucanases of the invention can hydrolyze the internal β-1, 4-glycosidic bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases of the invention also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, oil drilling and well stimulation, and in the fruit juice and brewing industry for the clarification and extraction of juices.

Inteins

In one aspect, the invention provides inteins, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In another aspect, the invention provides a chimeric protein comprising at least three domains, wherein the first domain comprises at least one enzyme domain or a binding protein domain, the second domain comprises at least one intein domain and a third domain comprising a detectable moiety domain, at least one intein domain is positioned between at least one enzyme or binding protein and at least one detectable moiety domain, and the intein domain has at least one cleavage or splicing activity.

In one aspect, the detectable moiety domain comprises a detectable peptide or polypeptide. The detectable peptide or a polypeptide can be a fluorescent peptide or polypeptide. The detectable peptide or a polypeptide can be a bioluminescent or a chemiluminescent peptide or polypeptide. In one aspect, the bioluminescent or chemiluminescent polypeptide comprises a green fluorescent protein (GFP), an aequorin, an obelin, a mnemiopsin or a berovin. In one aspect, the detectable moiety domain comprises an enzyme that generates a detectable signal. The enzyme that generates a detectable signal can comprise an alpha-galactosidase, an antibiotic (e.g., chloramphenicol acetyltransferase) or a kinase. The detectable moiety domain can comprise a radioactive isotope.

In one aspect, the chimeric protein is a recombinant fusion protein. In one aspect, the intein domain splicing activity results in cleavage of the enzyme domain from the intein domain and detectable domain. The intein domain splicing activity can result in cleavage of the enzyme domain from the intein domain and detectable domain and cleavage of the detectable domain from the intein domain. In one aspect, the intein domain splicing activity results in cleavage of the detectable domain from the intein domain. In one aspect, the intein domain has only splicing activity. The intein domain can have only cleaving activity.

In one aspect, at least one domain is separated from another domain by a linker. The linker can be a flexible linker. The intein domain can be separated from the detectable moiety domain and the enzyme domain by a linker.

Isomerases

In one aspect, the invention provides isomerases, e.g. xylose isomerases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having an isomerase activity, e.g. xylose isomerase activity, including thermostable and

thermotolerant isomerase activity, e.g. xylose isomerase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In one aspect, the invention provides xylose isomerase enzymes, polynucleotides encoding the enzymes, methods of making and using these polynucleotides and polypeptides. The polypeptides of the invention can be used in a variety of agricultural and industrial contexts. For example, the polypeptides of the invention can be used for converting glucose to fructose or for manufacturing high content fructose syrups in large quantities. Other examples include use of the polypeptides of the invention in confectionary, brewing, alcohol and soft drinks production, and in diabetic foods and sweeteners.

Laccases

In one aspect, the invention provides laccases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a laccase activity, including thermostable and thermotolerant laccase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In one aspect, the invention provides methods of depolymerizing lignin, e.g., in a pulp or paper manufacturing process, using a polypeptide of the invention. In another aspect, the invention provides methods for oxidizing products that can be mediators of laccase-catalyzed oxidation reactions, e.g., 2,2-azinobis-(3-ethylbenzthiazoline-6-sulfonate) (ABTS), 1- hydroxybenzotriazole (HBT), 2,2,6,6-tetramethylpiperidin-1-yloxy (TEMPO), dimethoxyphenol, dihydroxyfumaric acid (DHF) and the like.

Laccases are a subclass of the multicopper oxidase super family of enzymes, which includes ascorbate oxidases and the mammalian protein, ceruloplasmin. Laccases are one of the oldest known enzymes and were first implicated in the oxidation of urushiol and laccol. In one aspect, reactions catalyzed by laccases of the invention comprises the oxidation of phenolic substrates. The major target application has been in the delignification of wood fibers during the preparation of pulp.

Lipases

In one aspect, the invention provides lipases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a lipase activity, including

thermostable and thermotolerant lipase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

In one aspect, the lipases of the invention can be used in a variety of pharmaceutical, agricultural and industrial contexts, including the manufacture of cosmetics and nutraceuticals. In one aspect, the lipases of the invention are used in the biocatalytic synthesis of structured lipids (lipids that contain a defined set of fatty acids distributed in a defined manner on the glycerol backbone), including cocoa butter alternatives (CBA), lipids containing poly-unsaturated fatty acids (PUFAs), diacylglycerides, e.g., 1,3-diacyl glycerides (DAGs), monoglycerides, e.g., 2-monoglycerides (MAGs) and triacylglycerides (TAGs). In one aspect, the polypeptides of the invention are used to modify oils, such as fish, animal and vegetable oils, and lipids, such as poly-unsaturated fatty acids. The lipases of the invention can modify oils by hydrolysis, alcoholysis, esterification, transesterification and/or interesterification. The methods of the invention use lipases with defined regio-specificity or defined chemoselectivity in biocatalytic synthetic reactions. In another aspect, the polypeptides of the invention are used to synthesize enantiomerically pure chiral products.

The invention provides lipase enzymes, polynucleotides encoding the enzymes, methods of making and using these polynucleotides and polypeptides. The polypeptides of the invention can be used in a variety of pharmaceutical, agricultural and industrial contexts, including the manufacture of cosmetics and nutraceuticals. In one aspect, the polypeptides of the invention are used in the biocatalytic synthesis of structured lipids (lipids that contain a defined set of fatty acids distributed in a defined manner on the glycerol backbone), including cocoa butter alternatives, poly-unsaturated fatty acids (PUFAs), 1,3-diacyl glycerides (DAGs), 2-monoglycerides (MAGs) and triacylglycerides (TAGs), such as 1,3-dipalmitoyl-2-oleoylglycerol (POP), 1,3-distearoyl-2-oleoylglycerol (SOS), 1-palmitoyl-2-oleoyl-3-stearoylglycerol (POS) or 1-oleoyl-2,3-dimyristoylglycerol (OMM), long chain polyunsaturated fatty acids such as arachidonic acid, docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA).

In one aspect, the invention provides synthesis (using lipases of the invention) of a triglyceride mixture composed of POS (Palmitic-Oleic-Stearic), POP (Palmitic-Oleic-Palmitic) and SOS (Stearic-Oleic-Stearic) from glycerol. This synthesis uses free fatty acids versus fatty acid esters. In one aspect, this reaction can be performed in one pot with sequential addition of fatty acids using crude glycerol and free fatty acids and fatty acid esters. In one aspect, stearate and palmitate are mixed together to generate mixtures

of DAGs. In one aspect, the diacylglycerides are subsequently acylated with oleate to give components of cocoa butter equivalents. In alternative aspects, the proportions of POS, POP and SOS can be varied according to: stearate to palmitate ratio; selectivity of enzyme for palmitate versus stearate; or enzyme enantioselectivity (could alter levels of POS/SOP). One-pot synthesis of cocoa butter equivalents or other cocoa butter alternatives is possible using this aspect of the invention.

In one aspect, lipases that exhibit regioselectivity and/or chemoselectivity are used in the structure synthesis of lipids or in the processing of lipids. Thus, the methods of the invention use lipases with defined regio-specificity or defined chemoselectivity (e.g., a fatty acid specificity) in a biocatalytic synthetic reaction. For example, the methods of the invention can use lipases with SN1, SN2 and/or SN3 regio-specificity, or combinations thereof. In one aspect, the methods of the invention use lipases that exhibit regioselectivity for the 2-position of a triacylglyceride (TAG). This SN2 regioselectivity can be used in the synthesis of a variety of structured lipids, e.g., triacylglycerides (TAGs), including 1,3-DAGs and components of cocoa butter.

The methods and compositions (lipases) of the invention can be used in the biocatalytic synthesis of structured lipids, and the production of nutraceuticals (e.g., polyunsaturated fatty acids and oils), various foods and food additives (e.g., emulsifiers, fat replacers, margarines and spreads), cosmetics (e.g., emulsifiers, creams), pharmaceuticals and drug delivery agents (e.g., liposomes, tablets, formulations), and animal feed additives (e.g., polyunsaturated fatty acids, such as linoleic acids) comprising lipids made by the structured synthesis methods of the invention or processed by the methods of the invention

In one aspect, lipases of the invention can act on fluorogenic fatty acid (FA) esters, e.g., umbelliferyl FA esters. In one aspect, profiles of FA specificities of lipases made or modified by the methods of the invention can be obtained by measuring their relative activities on a series of umbelliferyl FA esters, such as palmitate, stearate, oleate, laurate, PUFA, butyrate.

The methods and compositions (lipases) of the invention can be used to synthesize enantiomerically pure chiral products. In one aspect, the methods and compositions (lipases) of the invention can be used to prepare a D-amino acid and corresponding esters from a racemic mix. For example, D-aspartic acid can be prepared from racemic aspartic acid. In one aspect, optically active D-homophenylalanine and/or its esters are prepared. The enantioselectively synthesized D-homophenylalanine can be starting material for

many drugs, such as Enalapril, Lisinopril, and Quinapril, used in the treatment of hypertension and congestive heart failure. The D-aspartic acid and its derivatives made by the methods and compositions of the invention can be used in pharmaceuticals, e.g., for the inhibition of arginiosuccinate synthetase to prevent or treat sepsis or cytokine-induced systemic hypotension or as immunosuppressive agents. The D-aspartic acid and its derivatives made by the methods and compositions of the invention can be used as taste modifying compositions for foods, e.g., as sweeteners (e.g., ALITAME™). For example, the methods and compositions (lipases) of the invention can be used to synthesize an optical isomer S(+) of 2-(6-methoxy-2-naphthyl) propionic acid from a racemic (R,S) ester of 2-(6-methoxy-2-naphthyl) propionic acid.

In one aspect, the methods and compositions (lipases) of the invention can be used to for stereoselectively hydrolyzing racemic mixtures of esters of 2-substituted acids, e.g., 2-aryloxy substituted acids, such as R-2-(4-hydroxyphenoxy)propionic acid, 2-arylpropionic acid, ketoprofen to synthesize enantiomerically pure chiral products.

The methods and compositions (lipases) of the invention can be used to hydrolyze oils, such as fish, animal and vegetable oils, and lipids, such as poly-unsaturated fatty acids. In one aspect, the polypeptides of the invention are used process fatty acids (such as poly-unsaturated fatty acids), e.g., fish oil fatty acids, for use in or as a feed additive.

Addition of poly-unsaturated fatty acids PUFAs to feed for dairy cattle has been demonstrated to result in improved fertility and milk yields. Fish oil contains a high level of PUFAs and therefore is a potentially inexpensive source for PUFAs as a starting material for the methods of the invention. The biocatalytic methods of the invention can process fish oil under mild conditions, thus avoiding harsh conditions utilized in some processes. Harsh conditions may promote unwanted isomerization, polymerization and oxidation of the PUFAs. In one aspect, the methods of the invention comprise lipase-catalyzed total hydrolysis of fish-oil or selective hydrolysis of PUFAs from fish oil to provide a mild alternative that would leave the high-value PUFAs intact. In one aspect, the methods further comprise hydrolysis of lipids by chemical or physical splitting of the fat.

In one aspect, the lipases and methods of the invention are used for the total hydrolysis of fish oil. Lipases can be screened for their ability to catalyze the total hydrolysis of fish oil under different conditions using. In alternative aspects, a single or multiple lipases are used to catalyze the total splitting of the fish oil. Several lipases of the invention may need to be used, owing to the presence of the PUFAs. In one aspect, a

PUFA-specific lipase of the invention is combined with a general lipase to achieve the desired effect.

The methods and compositions (lipases) of the invention can be used to catalyze the partial or total hydrolysis of other oils, e.g. olive oils, that do not contain PUFAs.

The methods and compositions (lipases) of the invention can be used to catalyze the hydrolysis of PUFA glycerol esters. These methods can be used to make feed additives. In one aspect, lipases of the invention catalyze the release of PUFAs from simple esters and fish oil. Standard assays and analytical methods can be utilized.

The methods and compositions (lipases) of the invention can be used to selectively hydrolyze saturated esters over unsaturated esters into acids or alcohols. The methods and compositions (lipases) of the invention can be used to treat latexes for a variety of purposes, e.g., to treat latexes used in hair fixative compositions to remove unpleasant odors. The methods and compositions (lipases) of the invention can be used in the treatment of a lipase deficiency in an animal, e.g., a mammal, such as a human. The methods and compositions (lipases) of the invention can be used to prepare lubricants, such as hydraulic oils. The methods and compositions (lipases) of the invention can be used in making and using detergents. The methods and compositions (lipases) of the invention can be used in processes for the chemical finishing of fabrics, fibers or yarns. In one aspect, the methods and compositions (lipases) of the invention can be used for obtaining flame retardancy in a fabric using, e.g., a halogen-substituted carboxylic acid or an ester thereof, i.e. a fluorinated, chlorinated or bromated carboxylic acid or an ester thereof.

Monooxygenases

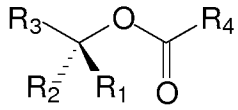
In one aspect, the invention provides monooxygenases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a monooxygenase activity, including thermostable and thermotolerant monooxygenase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

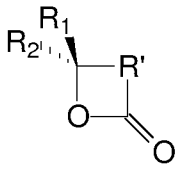
In one aspect, the monooxygenases of the invention have commercial utility as biocatalysts for use in the synthesis of aromatic and aliphatic esters and their derivatives, such as acids and alcohols. In one aspect, the monooxygenases of the invention are used in the catalysis of sulfoxidation reactions. In one aspect, the invention provides Baeyer-Villiger monooxygenases, polynucleotides encoding the Baeyer-Villiger

monooxygenases, and methods of using these Baeyer-Villiger monooxygenases and polynucleotides. In one aspect, the invention provides methods of producing chiral synthetic intermediates using Baeyer-Villiger monooxygenases.

In one aspect, the monooxygenase activity comprises catalysis of sulfoxidation reactions. The monooxygenase activity can comprise an asymmetric sulfoxidation reaction. The monooxygenase activity can be enantiospecific. In one aspect, it can generate a substantially chiral product.

In one aspect, the monooxygenase activity comprises generation of an ester or a lactone having at least one of the following structures:




wherein: $R_1$, $R_2$, $R_3$ and $R_4$ are each independently selected from -H, substituted or unsubstituted alkyl, alkenyl, alkynyl, aryl, heteroaryl, cycloalkyl, and heterocyclic; wherein the substituted groups are substituted with one or more of lower alkyl, hydroxy, alkoxy, mercapto, cycloalkyl, heterocyclic, aryl, heteroaryl, aryloxy, and halogen, or two or more of $R_1$, $R_2$, $R_3$ and $R_4$ may together form cyclic moieties, and, R′ is selected from substituted or unsubstituted alkylene, alkenylene, alkynylene, arylene, heteroarylene, cycloalkylene, and heterocyclic; wherein the substitutions are substituted with one or more of lower alkyl, hydroxy, alkoxy, mercapto, cycloalkyl, heterocyclic, aryl, heteroaryl, aryloxy, and halogen.

In one aspect, the monooxygenase activity comprises oxidation of a cycloalkanone to produce a chiral lactone. The cycloalkanone can comprise a cyclobutanone, a cyclopentanone, a cyclohexanone, a 2-methylcyclopentanone, a 2-methylcyclohexanone, a cyclohex-2-ene-1-one, a 2-(cyclohex-1-enyl)cyclohexanone, a 1,2-cyclohexanedione, a 1,3-cyclohexanedione or a 1,4-cyclohexanedione.

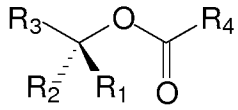
In one aspect, the monooxygenase activity comprises a chlorophenol 4-monooxygenase activity or a xylene monooxygenase activity.

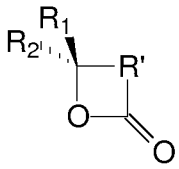
The invention provides a pharmaceutical composition comprising a polypeptide of the invention.

The invention provides a method for converting a ketone to its corresponding ester comprising contacting the ketone with a polypeptide of the invention under conditions wherein the polypeptide catalyzes the conversion of the ketone to its corresponding ester.

In one aspect, the polypeptide has an monooxygenase activity that is enantiospecific to generate a substantially chiral product. In one aspect, the ester is an aromatic or an aliphatic ester.

The invention provides a method for converting a cycloaliphatic ketone to its corresponding lactone comprising contacting the cycloaliphatic ketone with a polypeptide of the invention under conditions wherein the polypeptide catalyzes the conversion of the cycloaliphatic ketone to its corresponding lactone. In one aspect, the polypeptide has an monooxygenase activity that is enantiospecific to generate a substantially chiral product. In one aspect, the ester or lactone has at least one of the following structures:




wherein: $R_1$, $R_2$, $R_3$ and $R_4$ are each independently selected from -H, substituted or unsubstituted alkyl, alkenyl, alkynyl, aryl, heteroaryl, cycloalkyl, and heterocyclic; wherein the substituted groups are substituted with one or more of lower alkyl, hydroxy, alkoxy, mercapto, cycloalkyl, heterocyclic, aryl, heteroaryl, aryloxy, and halogen, or two or more of $R_1$, $R_2$, $R_3$ and $R_4$ may together form cyclic moieties, and, R′ is selected from substituted or unsubstituted alkylene, alkenylene, alkynylene, arylene, heteroarylene, cycloalkylene, and heterocyclic; wherein the substitutions are substituted with one or more of lower alkyl, hydroxy, alkoxy, mercapto, cycloalkyl, heterocyclic, aryl, heteroaryl, aryloxy, and halogen.

Nitroreductases

In one aspect, the invention provides nitroreductases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a nitroreductase activity, including thermostable and thermotolerant nitroreductase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Nitroreductases can catalyze the six-electron reduction of nitro compounds to the corresponding amines. Amines have a variety of applications as synthons and advanced pharmaceutical intermediates. There are markets for both aromatic amines and chiral aliphatic amines.

Nitroreductases of the invention fall in to two main classes. These are the oxygen-sensitive and oxygen-insensitive nitroreductases. The oxygen-sensitive enzyme can catalyze nitroreduction only under anaerobic conditions. A nitro anion radical is formed by a one-electron transfer and is immediately reoxidized in the presence of oxygen thus generating a futile cycle whereby reducing equivalents are consumed without nitroreduction. On the other hand the oxygen-insensitive nitroreductases catalyze nitroreduction in a series of two electron transfers, first via the nitroso and then the hydroxylamine intermediates before forming the amine.

Nitrilases

In one aspect, the invention provides nitrilases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a nitrilase activity, including thermostable and thermotolerant nitrilase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Nitrilases of the invention can be used for hydrolyzing a nitrile to a carboxylic acid. In one embodiment, the conditions of the reaction comprise aqueous conditions. In another embodiment, the conditions comprise a pH of about 8.0 and/or a temperature from about 37° C to about 45° C. Nitrilases of the invention can also be used for hydrolyzing a cyanohydrin moiety or an aminonitrile moiety of a molecule. Alternatively, the nitrilases of the invention can be used for making a chiral α-hydroxy acid molecule, a chiral amino acid molecule, a chiral β-hydroxy acid molecule, or a chiral gamma-hydroxy acid molecule. In one embodiment, the chiral molecule is an *(R)*-enantiomer. In another embodiment, the chiral molecule is an *(S)*-enantiomer. In one embodiment of the invention, one particular enzyme can have R-specificity for one particular substrate and the same enzyme can have S-specificity for a different particular substrate.

In one aspect, nitrilases of the invention can be used for making a composition or an intermediate thereof, wherein the nitrilase of the invention hydrolyzes a cyanohydrin or a aminonitrile moiety. In one embodiment, the composition or intermediate thereof comprises *(S)*-2-amino-4-phenyl butanoic acid. In a further embodiment, the composition or intermediate thereof comprises an L-amino acid. In a further embodiment, the composition comprises a food additive or a pharmaceutical drug.

In another aspect, nitrilases of the invention can be used for making an *(R)*-ethyl 4-cyano-3-hydroxybutyric acid, wherein the nitrilase of the invention acts upon a

hydroxyglutaryl nitrile and selectively produces an *(R)*-enantiomer, so as to make *(R)*-ethyl 4-cyano-3-hydroxybutyric acid. In one embodiment, the *ee* is at least 95% or at least 99%. In another embodiment, the hydroxyglutaryl nitrile comprises 1,3-di-cyano-2-hydroxy-propane or 3-hydroxyglutaronitrile.

In another aspect, nitrilases of the invention can be used for making an *(S)*-ethyl 4-cyano-3-hydroxybutyric acid, wherein the nitrilase of the invention acts upon a hydroxyglutaryl nitrile and selectively produces an *(S)*-enantiomer, so as to make *(S)*-ethyl 4-cyano-3-hydroxybutyric acid.

In another aspect, the nitrilases of the invention can be used for making a *(R)*-mandelic acid, wherein the nitrilase of the invention acts upon a mandelonitrile to produce a *(R)*-mandelic acid. In one embodiment, the *(R)*-mandelic acid comprises *(R)*-2-chloromandelic acid. In another embodiment, the *(R)*-mandelic acid comprises an aromatic ring substitution in the *ortho-*, *meta-*, or *para-* positions; a 1-naphthyl derivative of *(R)*-mandelic acid, a pyridyl derivative of *(R)*-mandelic acid or a thienyl derivative of *(R)*-mandelic acid or a combination thereof.

In another aspect, the nitrilases of the invention can be used for making a *(S)*-mandelic acid, wherein the nitrilase of the invention acts upon a mandelonitrile to produce a *(S)*-mandelic acid. In one embodiment, the *(S)*-mandelic acid comprises *(S)*-methyl benzyl cyanide and the mandelonitrile comprises *(S)*-methoxy-benzyl cyanide. In one embodiment, the *(S)*-mandelic acid comprises an aromatic ring substitution in the *ortho-*, *meta-*, or *para-* positions; a 1-naphthyl derivative of *(S)*-mandelic acid, a pyridyl derivative of *(S)*-mandelic acid or a thienyl derivative of *(S)*-mandelic acid or a combination thereof.

In yet another aspect, the nitrilases of the invention can be used for making a *(S)*-phenyl lactic acid derivative or a *(R)*-phenyllactic acid derivative, wherein the nitrilase of the invention acts upon a phenyllactonitrile and selectively produces an *(S)*-enantiomer or an *(R)*-enantiomer, thereby producing an *(S)*-phenyl lactic acid derivative or an *(R)*-phenyl lactic acid derivative.

P450 enzymes

In one aspect, the invention provides P450 enzymes, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a P450 enzymatic activity, including thermostable and thermotolerant P450 enzymatic activity, and

polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

P450s are oxidative enzymes that are widespread in nature and polypeptides of the invention having P450 activity can be used in processes such as detoxifying xenobiotics, catabolism of unusual carbon sources and biosynthesis of secondary metabolites (e.g., detoxification of toxic composition, e.g., pesticides, poisons, chemical warfare agents and the like). These oxygenases activate molecular oxygen using an iron-heme center and utilize a redox electron shuttle to support the epoxidation reaction.

In one aspect, the P450 activity comprises a monooxygenation reaction. In one aspect, the P450 activity comprises catalysis of incorporation of oxygen into a substrate. In one aspect, the P450 activity can further comprise hydroxylation of aliphatic or aromatic carbons. In another aspect, the P450 activity can comprise epoxidation. Alternatively, the P450 activity can comprise N-, O-, or S-dealkylation. In one aspect, the P450 activity can comprise dehalogenation. In another aspect the P450 activity can comprise oxidative deamination. Alternatively, the P450 activity can comprise N-oxidation or N-hydroxylation. In one aspect, the P450 activity can comprise sulphoxide formation.

In one aspect, the epoxidase activity further comprises an alkene substrate. The epoxidase activity can further comprise production of a chiral product. In one aspect, the epoxidase activity can be enantioselective.

Pectate lyases

In one aspect, the invention provides pectate lyases, e.g. pectinases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a pectate lyase, e.g. a pectinase activity, including thermostable and thermotolerant pectate lyase, e.g. a pectinase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

The pectate lyases, e.g. pectinases, of the invention can be used to catalyze the beta-elimination or hydrolysis of pectin and/or polygalacturonic acid, such as 1,4-linked alpha-D-galacturonic acid. They can be used in variety of industrial applications, e.g., to treat plant cell walls, such as those in cotton or other natural fibers. In another exemplary industrial application, the polypeptides of the invention can be used in textile scouring.

In one aspect, pectate lyase activity comprises catalysis of beta-elimination (trans-elimination) or hydrolysis of pectin or polygalacturonic acid (pectate). The pectate

lyase activity can comprise the breakup or dissolution of plant cell walls. The pectate lyase activity can comprise beta-elimination (trans-elimination) or hydrolysis of 1,4-linked alpha-D-galacturonic acid. The pectate lyase activity can comprise catalysis of beta-elimination (trans-elimination) or hydrolysis of methyl-esterified galacturonic acid. The pectate lyase activity can be exo-acting or endo-acting. In one aspect, the pectate lyase activity is endo-acting and acts at random sites within a polymer chain to give a mixture of oligomers. In one aspect, the pectate lyase activity is exo-acting and acts from one end of a polymer chain and produces monomers or dimers. The pectate lyase activity can catalyze the random cleavage of alpha-1,4-glycosidic linkages in pectic acid (polygalacturonic acid) by trans-elimination or hydrolysis. The pectate lyase activity can comprise activity the same or similar to pectate lyase (EC 4.2.2.2), poly(1,4-alpha-D-galacturonide) lyase, polygalacturonate lyase (EC 4.2.2.2), pectin lyase (EC 4.2.2.10), polygalacturonase (EC 3.2.1.15), exo-polygalacturonase (EC 3.2.1.67), exo-polygalacturonate lyase (EC 4.2.2.9) or exo-poly-alpha-galacturonosidase (EC 3.2.1.82). The pectate lyase activity can comprise beta-elimination (trans-elimination) or hydrolysis of galactan to galactose or galactooligomers. The pectate lyase activity can comprise beta-elimination (trans-elimination) or hydrolysis of a plant fiber. The plant fiber can comprise cotton fiber, hemp fiber or flax fiber.

The pectate lyases, e.g. pectinases, of the invention can be used for hydrolyzing, breaking up or disrupting a pectin- or pectate (polygalacturonic acid)-comprising composition, for liquefying or removing a pectin or pectate (polygalacturonic acid) from a composition. Alternatively, the pectate lyases, e.g. pectinases, of the invention can be used in detergent compositions. In one aspect, the pectate lyase is a nonsurface-active pectate lyase or a surface-active pectate lyase. The pectate lyase can be formulated in a non-aqueous liquid composition, a cast solid, a granular form, a particulate form, a compressed tablet, a gel form, a paste or a slurry form.

In one aspect, the pectate lyases, e.g. pectinases, of the invention can be used for washing an object. In another aspect, textiles or fabrics comprise a polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention, wherein the polypeptide has pectate lyase, e.g. pectinase activity. Additionally, the pectate lyases, e.g. pectinases, of the invention can be used for fiber, thread, textile or fabric scouring. In one aspect, the pectate lyase is an alkaline active and thermostable pectate lyase. The desizing and scouring treatments can be combined in a single bath. The method can further comprise addition of an alkaline and thermostable amylase. The desizing or

scouring treatments can comprise conditions of between about pH 8.5 to pH 10.0 and temperatures of at about 40°C. The method can further comprise addition of a bleaching step. The desizing, scouring and bleaching treatments can be done simultaneously or sequentially in a single-bath container. The bleaching treatment can comprise hydrogen peroxide or at least one peroxy compound that can generate hydrogen peroxide when dissolved in water, or combinations thereof, and at least one bleach activator. The fiber, thread, textile or fabric can comprise a cellulosic material. The cellulosic material can comprise a crude fiber, a yarn, a woven or knit textile, a cotton, a linen, a flax, a ramie, a rayon, a hemp, a jute or a blend of natural or synthetic fibers.

Alternatively, the pectate lyases, e.g. pectinases, of the invention can be used in feeds or foods. For example, the pectate lyases, e.g. pectinases, of the invention can be used to improve the extraction of oil from an oil-rich plant material. In one aspect, the oil-rich plant material comprises an oil-rich seed. The oil can be a soybean oil, an olive oil, a rapeseed (canola) oil or a sunflower oil.

In another aspect, the pectate lyases, e.g. pectinases, of the invention can be used for preparing a fruit or vegetable juice, syrup, puree or extract. In yet another aspect, the pectate lyases, e.g. pectinases, of the invention can used for treating a paper or a paper or wood pulp. Alternatively, the invention provides papers or paper products or paper pulps comprising a pectate lyase of the invention, or a polypeptide encoded by a nucleic acid of the invention.

In yet another aspect, the invention provides pharmaceutical compositions comprising a polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention, wherein the polypeptide has pectate lyase, e.g. pectinase activity. The pharmaceutical composition can act as a digestive aid.

Alternatively, the invention provides oral care products comprising a polypeptide of the invention, or a polypeptide encoded by a nucleic acid of the invention, wherein the polypeptide has pectate lyase, e.g. pectinase activity. The oral care product can comprise a toothpaste, a dental cream, a gel or a tooth powder, an odontic, a mouth wash, a pre- or post brushing rinse formulation, a chewing gum, a lozenge or a candy.

Phosphatases

In one aspect, the invention provides phosphatases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a phosphatase activity, including thermostable and thermotolerant phosphatase activity, and

polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Phosphatases are a group of enzymes that remove phosphate groups from organophosphate ester compounds. There are numerous phosphatases, including alkaline phosphatases, phosphodiesterases and phytases.

Alkaline phosphatases are widely distributed enzymes and are composed of a group of enzymes which hydrolyze organic phosphate ester bonds at alkaline pH.

Phosphodiesterases are capable of hydrolyzing nucleic acids by hydrolyzing the phosphodiester bridges of DNA and RNA. The classification of phosphodiesterases depends upon which side of the phosphodiester bridge is attacked. The 3' enzymes specifically hydrolyze the ester linkage between the 3' carbon and the phosphoric group whereas the 5' enzymes hydrolyze the ester linkage between the phosphoric group and the 5' carbon of the phosphodiester bridge. The best known of the class 3' enzymes is a phosphodiesterase from the venom of the rattlesnake or from a rustle's viper, which hydrolyses all the 3' bonds in either RNA or DNA liberating nearly all the nucleotide units as nucleotide 5' phosphates. This enzyme requires a free 3' hydroxyl group on the terminal nucleotide residue and proceeds stepwise from that end of the polynucleotide chain. This enzyme and all other nucleases which attack only at the ends of the polynucleotide chains are called exonucleases. The 5' enzymes are represented by a phosphodiesterase from bovine spleen, also an exonuclease, which hydrolyses all the 5' linkages of both DNA and RNA and thus liberates only nucleoside 3' phosphates. It begins its attack at the end of the chain having a free 3' hydroxyl group.

Phytase enzymes remove phosphate from phytic acid (inositol hexaphosphoric acid), a compound found in plants such as corn, wheat and rice. The enzyme has commercial use for the treatment of animal feed, making the inositol of the phytic acid available for animal nutrition. Phytases are used to improve the utilization of natural phosphorus in animal feed. Use of phytase as a feed additive enables the animal to metabolize a larger degree of its cereal feed's natural mineral content thereby reducing or altogether eliminating the need for synthetic phosphorus additives. More important than the reduced need for phosphorus additives is the corresponding reduction of phosphorus in pig and chicken waste. Many European countries severely limit the amount of manure that can be spread per acre due to concerns regarding phosphorus contamination of ground water.

Alkaline phosphatases hydrolyze monophosphate esters, releasing an organic phosphate and the cognate alcohol compound. It is non-specific with respect to the alcohol moiety and it is this feature which accounts for the many uses of this enzyme.

Phospholipases

In one aspect, the invention provides phospholipases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a phospholipase activity, including thermostable and thermotolerant phospholipase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Phospholipases are enzymes that hydrolyze the ester bonds of phospholipids. Corresponding to their importance in the metabolism of phospholipids, these enzymes are widespread among prokaryotes and eukaryotes. The phospholipases affect the metabolism, construction and reorganization of biological membranes and are involved in signal cascades. Several types of phospholipases are known which differ in their specificity according to the position of the bond attacked in the phospholipid molecule. Phospholipase A1 (PLA1) removes the 1-position fatty acid to produce free fatty acid and 1-lyso-2-acylphospholipid. Phospholipase A2 (PLA2) removes the 2-position fatty acid to produce free fatty acid and 1-acyl-2-lysophospholipid. PLA1 and PLA2 enzymes can be intra- or extra-cellular, membrane-bound or soluble. Intracellular PLA2 is found in almost every mammalian cell. Phospholipase C (PLC) removes the phosphate moiety to produce 1,2 diacylglycerol and phospho base. Phospholipase D (PLD) produces 1,2-diacylglycerophosphate and base group. PLC and PLD are important in cell function and signaling. PLD had been the dominant phospholipase in biocatalysis. Patatins are another type of phospholipase, thought to work as a PLA.

The invention provides methods for cleaving a glycerolphosphate ester linkage comprising the following steps: (a) providing a polypeptide having a phospholipase activity, wherein the polypeptide comprises an amino acid sequence of the invention, or the polypeptide is encoded by a nucleic acid of the invention; (b) providing a composition comprising a glycerolphosphate ester linkage; and, (c) contacting the polypeptide of step (a) with the composition of step (b) under conditions wherein the polypeptide cleaves the glycerolphosphate ester linkage. In one aspect, the conditions comprise between about pH 5 to about 5.5, or, between about pH 4.5 to about 5.0. In one aspect, the conditions comprise a temperature of between about 40°C and about 70°C. In

one aspect, the composition comprises a vegetable oil. In one aspect, the composition comprises an oilseed phospholipid. In one aspect, the cleavage reaction can generate a water extractable phosphorylated base and a diglyceride.

Phospholipases of the invention can be used in oil degumming, wherein the phospholipase is used under conditions wherein the phospholipase can cleave ester linkages in an oil, thereby degumming the oil. In one aspect, the oil is a vegetable oil . In another aspect, the vegetable oil comprises oilseed. The vegetable oil can comprise palm oil, rapeseed oil, corn oil, soybean oil, canola oil, sesame oil, peanut oil or sunflower oil. In one aspect, the method further comprises addition of a phospholipase of the invention, another phospholipase, another enzyme, or a combination thereof.

In another aspect of the invention, phospholipases of the invention can be used for converting a non-hydratable phospholipid to a hydratable form or for caustic refining of a phospholipid-containing composition. In the latter use, the polypeptide of the invention can be added before caustic refining and the composition comprising the phospholipid can comprise a plant and the polypeptide can be expressed transgenically in the plant, the polypeptide having a phospholipase activity can be added during crushing of a seed or other plant part, or, the polypeptide having a phospholipase activity is added following crushing or prior to refining. The polypeptide can be added during caustic refining and varying levels of acid and caustic can be added depending on levels of phosphorous and levels of free fatty acids. The polypeptide can be added after caustic refining: in an intense mixer or retention mixer prior to separation; following a heating step; in a centrifuge; in a soapstock; in a washwater; or, during bleaching or deodorizing steps.

In yet another aspect, the phospholipases of the invention can be used for purification of a phytosterol or a triterpene. The phytosterol or a triterpene can comprise a plant sterol. The plant sterol can be derived from a vegetable oil. The vegetable oil can comprise a coconut oil, canola oil, cocoa butter oil, corn oil, cottonseed oil, linseed oil, olive oil, palm oil, peanut oil, oil derived from a rice bran, safflower oil, sesame oil, soybean oil or a sunflower oil. The method can comprise use of nonpolar solvents to quantitatively extract free phytosterols and phytosteryl fatty-acid esters. The phytosterol or a triterpene can comprise a β-sitosterol, a campesterol, a stigmasterol, a stigmastanol, a β-sitostanol, a sitostanol, a desmosterol, a chalinasterol, a poriferasterol, a clionasterol or a brassicasterol.

In one embodiment, the phospholipases of the invention can be used for refining a crude oil. The polypeptide can have a phospholipase activity is in a water solution that is

added to the composition. The water level can be between about 0.5 to 5%. The process time can be less than about 2 hours, less than about 60 minutes, less than about 30 minutes, less than 15 minutes, or less than 5 minutes. The hydrolysis conditions can comprise a temperature of between about 25°C-70°C. The hydrolysis conditions can comprise use of caustics. The hydrolysis conditions can comprise a pH of between about pH 3 and pH 10, between about pH 4 and pH 9, or between about pH 5 and pH 8. The hydrolysis conditions can comprise addition of emulsifiers and/or mixing after the contacting of step (c). The methods can comprise addition of an emulsion-breaker and/or heat to promote separation of an aqueous phase. The methods can comprise degumming before the contacting step to collect lecithin by centrifugation and then adding a PLC, a PLC and/or a PLA to remove non-hydratable phospholipids. The methods can comprise water degumming of crude oil to less than 10 ppm for edible oils and subsequent physical refining to less than about 50 ppm for biodiesel oils. The methods can comprise addition of acid to promote hydration of non-hydratable phospholipids.

Phytases

In one aspect, the invention provides phytases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a phytase activity, including thermostable and thermotolerant phytase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Conversion of phytate to inositol and inorganic phosphorous can be catalyzed by phytase enzymes. Phytases such as phytase #EC 3.1.3.8 are capable of catalyzing the hydrolysis of myo-inositol hexaphosphate to D-myo-inositol 1,2,4,5,6-pentaphosphate and orthophosphate. Other phytases hydrolyze inositol pentaphosphate to tetra-, tri-, and lower phosphates. Acid phosphatases are enzymes that catalytically hydrolyze a wide variety of phosphate esters. For example, #EC 3.1.3.2 enzymes catalyze the hydrolysis of orthophosphoric monoesters to orthophosphate products.

Phytases of the invention can be used in producing phytase as a feed additive, e.g. for monogastric animals, fish, poultry, ruminants and other non-ruminants. Phytases of the invention can also be used for producing animal feed from certain industrial processes, *e.g.*, wheat and corn waste products. In one aspect, the wet milling process of corn produces glutens sold as animal feeds. The addition of phytase improves the nutritional value of the feed product.

Phytases of the invention may also be used in dietary aids or in pharmaceutical compositions, for reducing pollution and increasing nutrient availability in an environment or environmental sample by degrading environmental phytic acid, for liberating minerals from phytates in plant materials either *in vitro*, *i.e.*, in feed treatment processes, or *in vivo*, *i.e.*, by administering the enzymes to animals.

Polymerases

In one aspect, the invention provides polymerases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a polymerase activity, including thermostable and thermotolerant polymerase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

The polymerase enzymes of the invention can have different polymerase activities at various high temperatures. In one aspect, the polymerase activity comprises addition of deoxynucleotides at the 3' hydroxyl end of a polynucleotide. The invention also provides kits, e.g., diagnostic kits, and methods for performing various amplification reactions, e.g., polymerase chain reactions, transcription amplifications, ligase chain reactions, self-sustained sequence replication or Q Beta replicase amplifications.

In one aspect, the polymerase activity comprises addition of nucleotides at the 3' hydroxyl end of a nucleic acid. The polymerase activity can comprise a 5'→3' polymerase activity, a 3'→5' exonuclease activity or a 5'→3' exonuclease activity or all or a combination thereof. In one aspect, the polymerase activity comprises only a 5'→3' polymerase activity, but not a 3'→5' exonuclease activity or a 5'→3' exonuclease activity. In another aspect, the polymerase activity can comprise a 5'→3' polymerase activity and a 3'→5' exonuclease activity, but not a 5'→3' exonuclease activity. Alternatively, the polymerase activity can comprise a 5'→3' polymerase activity and a 5'→3' exonuclease activity, but not a 3'→5' exonuclease activity. The polymerase activity can comprise addition of dUTP or dITP. The polymerase activity can comprise addition of a modified or a non-natural nucleotide to a polynucleotide, such as an analog of guanine, cytosine, thymine, adenine or uracil, e.g., a 2-aminopurine, an inosine or a 5-methylcytosine.

In one aspect, the polymerase activity can comprise strand displacement properties. In one aspect, the polymerase activity comprises reverse transcriptase activity.

Proteases

In one aspect, the invention provides proteases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a protease activity, including thermostable and thermotolerant protease activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Proteases of the invention can be carbonyl hydrolases which act to cleave peptide bonds of proteins or peptides. Proteolytic enzymes are ubiquitous in occurrence, found in all living organisms, and are essential for cell growth and differentiation. The extracellular proteases are of commercial value and find multiple applications in various industrial sectors. Industrial applications of proteases include food processing, brewing, alcohol production, peptide synthesis, enantioselectivity, hide preparation in the leather industry, waste management and animal degradation, silver recovery in the photographic industry, medical treatment, silk degumming, biofilm degradation, biomass conversion to a biofuel (e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel), biodefense, antimicrobial agents and disinfectants, personal care and cosmetics, biotech reagents and in increasing starch yield from corn wet milling. Additionally, proteases are important components of laundry detergents and other products. Within biological research, proteases are used in purification processes to degrade unwanted proteins. It is often desirable to employ proteases of low specificity or mixtures of more specific proteases to obtain the necessary degree of degradation.

Proteases are classified according to their catalytic mechanisms. The International Union of Biochemistry and Molecular Biology (IUBMB) recognizes four mechanistic classes: (1) the serine proteases; (2) the cysteine proteases; (3) the aspartic proteases; and (4) the metalloproteases. In addition, the IUBMB recognizes a class of endopeptidases (oligopeptidases) of unknown catalytic mechanism. The serine proteases have alkaline pH optima, the metalloproteases are optimally active around neutrality, and the cysteine and aspartic enzymes have acidic pH optima. Serine proteases class comprises two distinct families: the chymotrypsin family, which includes the mammalian enzymes such as chymotrypsin, trypsin, elastase, or kallikrein, and the subtilisin family, which include the bacterial enzymes such as subtilisin. Serine proteases are used for a variety of industrial purposes, such as laundry detergents to aid in the removal of proteinaceous stains. In the food processing industry, serine proteases are used to produce protein-rich concentrates from fish and livestock, and in the preparation of dairy products.

The proteases of the invention can be used in a variety of diagnostic, therapeutic, and industrial contexts. The proteases of the invention can be used as, e.g., an additive for a detergent, for processing foods and for chemical synthesis utilizing a reverse reaction. Additionally, the proteases of the invention can be used in food processing, brewing, bath additives, alcohol production, peptide synthesis, enantioselectivity, hide preparation in the leather industry, waste management and animal degradation, silver recovery in the photographic industry, medical treatment, silk degumming, biofilm degradation, biomass conversion to biofuel (e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel), biodefense, antimicrobial agents and disinfectants, personal care and cosmetics, biotech reagents, in increasing starch yield from corn wet milling and pharmaceuticals such as digestive aids and anti-inflammatory (anti-phlogistic) agents.

Xylanases

In one aspect, the invention provides xylanases, polynucleotides encoding them, and methods of making and using these polynucleotides and polypeptides. In one aspect, the invention is directed to polypeptides, e.g., enzymes, having a xylanase activity, including thermostable and thermotolerant xylanase activity, and polynucleotides encoding these enzymes, and making and using these polynucleotides and polypeptides.

Xylanases (e.g., endo-1,4-beta-xylanase, EC 3.2.1.8) of the invention can hydrolyze internal β-1,4-xylosidic linkages in xylan to produce smaller molecular weight xylose and xylo-oligomers. Xylans are polysaccharides formed from 1,4-β-glycoside-linked D-xylopyranoses. Xylanases of the invention are of considerable commercial value, being used in the food industry, for baking and fruit and vegetable processing, breakdown of agricultural waste, in the manufacture of animal feed and in pulp and paper production.

Arabinoxylanase are major non-starch polysaccharides of cereals representing 2.5 – 7.1% w/w depending on variety and growth conditions. The physicochemical properties of this polysaccharide are such that it gives rise to viscous solutions or even gels under oxidative conditions. In addition, arabinoxylans have high water-binding capacity and may have a role in protein foam stability. All of these characteristics present problems for several industries including brewing, baking, animal nutrition and paper manufacturing. In brewing applications, the presence of xylan results in wort filterability and haze formation issues. In baking applications (especially for cookies and crackers), these arabinoxylans create sticky doughs that are difficult to machine and reduce biscuit

size. In addition, this carbohydrate is implicated in rapid rehydration of the baked product resulting in loss of crispiness and reduced shelf-life. For monogastric animal feed applications with cereal diets, arabinoxylan is a major contributing factor to viscosity of gut contents and thereby adversely affects the digestibility of the feed and animal growth rate. For ruminant animals, these polysaccharides represent substantial components of fiber intake and more complete digestion of arabinoxylans would facilitate higher feed conversion efficiencies.

Xylanases are currently used as additives (dough conditioners) in dough processing for the hydrolysis of water soluble arabinoxylan. In baking applications (especially for cookies and crackers), arabinoxylan creates sticky doughs that are difficult to machine and reduce biscuit size. In addition, this carbohydrate is implicated in rapid rehydration of the baked product resulting in loss of crispiness and reduced shelf-life.

The enhancement of xylan digestion in animal feed may improve the availability and digestibility of valuable carbohydrate and protein feed nutrients. For monogastric animal feed applications with cereal diets, arabinoxylan is a major contributing factor to viscosity of gut contents and thereby adversely affects the digestibility of the feed and animal growth rate. For ruminant animals, these polysaccharides represent substantial components of fiber intake and more complete digestion would facilitate higher feed conversion efficiencies. It is desirable for animal feed xylanases to be active in the animal stomach. This requires a feed enzyme to have high activity at 37 °C and at low pH for monogastrics (pH 2-4) and near neutral pH for ruminants (pH 6.5-7). The enzyme should also possess resistance to animal gut xylanases and stability at the higher temperatures involved in feed pelleting. As such, there is a need in the art for xylanase feed additives for monogastric feed with high specific activity, activity at 35-40°C and pH 2-4, half life greater than 30 minutes in SGF and a half-life > 5 minutes at 85°C in formulated state. For ruminant feed, there is a need for xylanase feed additives that have a high specific activity, activity at 35-40°C and pH 6.5-7.0, half life greater than 30 minutes in SRF and stability as a concentrated dry powder.

In one aspect, the xylanases of the invention are also used in improving the quality and quantity of milk protein production in lactating cows, increasing the amount of soluble saccharides in the stomach and small intestine of pigs, improving late egg production efficiency and egg yields in hens. Additionally, xylanases of the inventions can be used in biobleaching and treatment of chemical pulps, biobleaching and treatment of wood or paper pulps, in reducing lignin in wood and modifying wood, as feed

additives and/or supplements or in manufacturing cellulose solutions. Detergent compositions comprising xylanases of the invention are used for fruit, vegetables and/or mud and clay compounds.

In another aspect, xylanases of the invention can be used in compositions for the treatments and/or prophylaxis of coccidiosis. In yet another aspect, xylanases of the invention can be used in the production of water soluble dietary fiber, in improving the filterability, separation and production of starch, the beverage industry in improving filterability of wort or beer, in reducing viscosity of plant material, or in increasing viscosity or gel strength of food products such as jam, marmalade, jelly, juice, paste, soup, salsa, etc. Xylanases of the invention may also be used in hydrolysis of hemicellulose for which it is selective, particularly in the presence of cellulose. In addition, xylanases of the invention can also be used in the production of bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel, in transformation of a microbe that produces ethanol, methanol, butanol, propanol, etc, in production of oenological tannins and enzymatic composition, in stimulating the natural defenses of plants, in production of sugars from hemicellulose substrates, in the cleaning of fruit, vegetables, mud or clay containing soils, in cleaning beer filtration membranes, and in killing or inhibiting microbial cells.

Tables 1 and 4 detail exemplary activities of polypeptides of the invention (see also Table 3); noting that each polypeptide of the invention can have more than one specific enzymatic activity. The activities/ functions of exemplary polypeptides of the invention were determined by sequence comparison (BLAST) analysis with public sequence databases, such as the NR database available through GenBank and the Geneseq database available from Thomson Scientific, as summarized in Table 1 and Table 4. Table 1 and Table 4 describe the source organism of the closest hit polypeptide (see "Organism" column); the GenBank accession number of the top BLAST hit for DNA and protein (see "NR or Geneseq Protein Accession Code"), the percent sequence identity between the sequence of the invention and the top BLAST hit (see "% Id" column), and other descriptions for that particular exemplary polynucleotide/polypeptide entry and the BLAST analysis. The EC (Enzyme Commission) Number, or enzyme nomenclature, is based upon the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB); see also Table 3, below.

For example, as an aid in reading Table 1 and Table 4, the polypeptide SEQ ID NO:16,334 (and active fragments thereof), encoded, e.g., by SEQ ID NO:16,333, has at

least a transaminase activity (including, in one aspect, a (2-aminoethyl)phosphonate-pyruvate transaminase activity), and its activity was determined by a closest BLAST hit from a sequence initially isolated from *Clostridium tetani* strain E88; NR or Geneseq Protein Accession Code 28203795; and SEQ ID NO:16,334/16,333 have at least about 58% sequence identity to this closest hit.

Table 1 and Table 4 list exemplary enzymatic activities of polypeptides of the invention, as can be determined by sequence identity (e.g., homology); and in one embodiment a sequence of the invention comprises an enzyme having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity (homology) to a polypeptide encoded by an exemplary nucleic acid sequence of the invention, including all odd numbered SEQ ID NO:1 to SEQ ID NO:108,699, or an exemplary polypeptide of the invention, including all even numbered SEQ ID NO:2 to SEQ ID NO:108,699, and with an exemplary function as listed in Table 1 and Table 4.

A polypeptide or peptide of the invention can have at least one activity as set forth in Table 1 and Table 4 or as listed in Table 3, below:

Table 3: EC (Enzyme Commission) Numbers with the corresponding mode of action for each enzyme class, subclass and sub-subclass

| | |
|---|---|
| 1. -. -.- Oxidoreductases. | 1. 6. 3.- With a oxygen as acceptor. |
| 1. 1. -.- Acting on the CH-OH group of donors. | 1. 6. 4.- With a disulfide as acceptor. |
| 1. 1. 1.- With NAD(+) or NADP(+) as acceptor. | 1. 6. 5.- With a quinone or similar compound as acceptor. |
| 1. 1. 2.- With a cytochrome as acceptor. | 1. 6. 6.- With a nitrogenous group as acceptor. |
| 1. 1. 3.- With oxygen as acceptor. | 1. 6. 8.- With a flavin as acceptor. |
| 1. 1. 4.- With a disulfide as acceptor. | 1. 6.99.- With other acceptors. |
| 1. 1. 5.- With a quinone or similar compound as acceptor. | 1. 7. -.- Acting on other nitrogenous compounds as donors. |
| 1. 1.99.- With other acceptors. | 1. 7. 1.- With NAD(+) or NADP(+) as acceptor. |
| 1. 2. -.- Acting on the aldehyde or oxo group of donors. | 1. 7. 2.- With a cytochrome as acceptor. |
| 1. 2. 1.- With NAD(+) or NADP(+) as acceptor. | 1. 7. 3.- With oxygen as acceptor. |
| 1. 2. 2.- With a cytochrome as acceptor. | 1. 7. 7.- With an iron-sulfur protein as acceptor. |
| 1. 2. 3.- With oxygen as acceptor. | 1. 7.99.- With other acceptors. |
| 1. 2. 4.- With a disulfide as acceptor. | 1. 8. -.- Acting on a sulfur group of donors. |
| 1. 2. 7.- With an iron-sulfur protein as acceptor. | 1. 8. 1.- With NAD(+) or NADP(+) as acceptor. |
| 1. 2.99.- With other acceptors. | 1. 8. 2.- With a cytochrome as acceptor. |
| 1. 3. -.- Acting on the CH-CH group of donors. | 1. 8. 3.- With oxygen as acceptor. |
| 1. 3. 1.- With NAD(+) or NADP(+) as acceptor. | 1. 8. 4.- With a disulfide as acceptor. |
| 1. 3. 2.- With a cytochrome as acceptor. | 1. 8. 5.- With a quinone or similar compound as acceptor. |

| | |
|---|---|
| 1. 3. 3.- With oxygen as acceptor. | 1. 8. 7.- With an iron-sulfur protein as acceptor. |
| 1. 3. 5.- With a quinone or related compound as acceptor. | 1. 8.98.- With other, known, acceptors. |
| 1. 3. 7.- With an iron-sulfur protein as acceptor. | 1. 8.99.- With other acceptors. |
| 1. 3.99.- With other acceptors. | 1. 9. -.- Acting on a heme group of donors. |
| 1. 4. -.- Acting on the CH-NH(2) group of donors. | 1. 9. 3.- With oxygen as acceptor. |
| 1. 4. 1.- With NAD(+) or NADP(+) as acceptor. | 1. 9. 6.- With a nitrogenous group as acceptor. |
| 1. 4. 2.- With a cytochrome as acceptor. | 1. 9.99.- With other acceptors. |
| 1. 4. 3.- With oxygen as acceptor. | 1.10. -.- Acting on diphenols and related substances as donors. |
| 1. 4. 4.- With a disulfide as acceptor. | 1.10. 1.- With NAD(+) or NADP(+) as acceptor. |
| 1. 4. 7.- With an iron-sulfur protein as acceptor. | 1.10. 2.- With a cytochrome as acceptor. |
| 1. 4.99.- With other acceptors. | 1.10. 3.- With oxygen as acceptor. |
| 1. 5. -.- Acting on the CH-NH group of donors. | 1.10.99.- With other acceptors. |
| 1. 5. 1.- With NAD(+) or NADP(+) as acceptor. | 1.11. -.- Acting on a peroxide as acceptor (peroxidases). |
| 1. 5. 3.- With oxygen as acceptor. | 1.12. -.- Acting on hydrogen as donor. |
| 1. 5. 4.- With a disulfide as acceptor. | 1.12. 1.- With NAD(+) or NADP(+) as acceptor. |
| 1. 5. 5.- With a quinone or similar compound as acceptor. | 1.12. 2.- With a cytochrome as acceptor. |
| 1. 5. 8.- With a flavin as acceptor. | 1.12. 5.- With a quinone or similar compound as acceptor. |
| 1. 5.99.- With other acceptors. | 1.12. 7.- With an iron-sulfur protein as acceptor. |
| 1. 6. -.- Acting on NADH or NADPH. | 1.12.98.- With other known acceptors. |
| 1. 6. 1.- With NAD(+) or NADP(+) as acceptor. | 1.12.99.- With other acceptors. |
| 1. 6. 2.- With a heme protein as acceptor. | 1.13. -.- Acting on single donors with incorporation of molecular oxygen. |

| | |
|---|---|
| 1.13.11.- With incorporation of two atoms of oxygen. | 1.19. 6.- With dinitrogen as acceptor. |
| 1.13.12.- With incorporation of one atom of oxygen. | 1.20. -.- Acting on phosphorus or arsenic in donors. |
| 1.14. -.- Acting on paired donors, with incorporation or reduction of molecular oxygen | 1.20. 1.- Acting on phosphorus or arsenic in donors, with NAD(P)(+) as acceptor |
| 1.14.11.- With 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors | 1.20. 4.- Acting on phosphorus or arsenic in donors, with disulfide as acceptor |
| 1.14.12.- With NADH or NADPH as one donor, and incorporation of two atoms of oxygen into one donor | 1.20.98.- Acting on phosphorus or arsenic in donors, with other, known acceptors |
| 1.14.13.- With NADH or NADPH as one donor, and incorporation of one atom of oxygen | 1.20.99.- Acting on phosphorus or arsenic in donors, with other acceptors |
| 1.14.14.- With reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen | 1.21. -.- Acting on x-H and y-H to form an x-y bond. |
| 1.14.15.- With a reduced iron-sulfur protein as one donor, and incorporation of one atom of oxygen | 1.21. 3.- With oxygen as acceptor. |
| 1.14.16.- With reduced pteridine as one donor, and incorporation of one atom of oxygen | 1.21. 4.- With a disulfide as acceptor. |
| 1.14.17.- With reduced ascorbate as one donor, and incorporation of one atom of oxygen | 1.21.99.- With other acceptors. |
| 1.14.18.- With another compound as one donor, and incorporation of one incorporation of one atom of oxygen | 1.97. -.- Other oxidoreductases. |
| 1.14.19.- With oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water | 2. -. -.- Transferases. |

| | |
|---|---|
| 1.14.20.- With 2-oxoglutarate as one donor, and the other dehydrogenated. | 2. 1. -.- Transferring one-carbon groups. |
| 1.14.21.- With NADH or NADPH as one donor, and the other dehydrogenated. | 2. 1. 1.- Methyltransferases. |
| 1.15. -.- Acting on superoxide as acceptor. | 2. 1. 2.- Hydroxymethyl-, formyl- and related transferases. |
| 1.16. -.- Oxidizing metal ions. | 2. 1. 3.- Carboxyl- and carbamoyltransferases. |
| 1.16. 1.- With NAD(+) or NADP(+) as acceptor. | 2. 1. 4.- Amidinotransferases. |
| 1.16. 3.- With oxygen as acceptor. | 2. 2. -.- Transferring aldehyde or ketone residues. |
| 1.16. 8.- With flavin as acceptor. | 2. 2. 1.- Transketolases and transaldolases. |
| 1.17. -.- Acting on CH or CH(2) groups. | 2. 3. -.- Acyltransferases. |
| 1.17. 1.- With NAD(+) or NADP(+) as acceptor. | 2. 3. 1.- Transferring groups other than amino-acyl groups. |
| 1.17. 3.- With oxygen as acceptor. | 2. 3. 2.- Aminoacyltransferases. |
| 1.17. 4.- With a disulfide as acceptor. | 2. 3. 3.- Acyl groups converted into alkyl on transfer. |
| 1.17. 5.- With a quinone or similar compound as acceptor | 2. 4. -.- Glycosyltransferases. |
| 1.17.99.- With other acceptors. | 2. 4. 1.- Hexosyltransferases. |
| 1.18. -.- Acting on iron-sulfur proteins as donors. | 2. 4. 2.- Pentosyltransferases. |
| 1.18. 1.- With NAD(+) or NADP(+) as acceptor. | 2. 4.99.- Transferring other glycosyl groups. |
| 1.18. 6.- With dinitrogen as acceptor. | 2. 5. -.- Transferring alkyl or aryl groups, other than methyl groups. |
| 1.18.96.- With other, known, acceptors. | 2. 6. -.- Transferring nitrogenous groups. |
| 1.18.99.- With H(+) as acceptor. | 2. 6. 1.- Transaminases (aminotransferases). |
| 1.19. -.- Acting on reduced flavodoxin as donor. | 2. 6. 3.- Oximinotransferases. |

| | |
|---|---|
| 2. 6.99.- Transferring other nitrogenous groups. | 3. 1.22.- Endodeoxyribonucleases producing other than 5'-phosphomonoesters. |
| 2. 7. -.- Transferring phosphorous-containing groups. | 3. 1.25.- Site-specific endodeoxyribonucleases specific for altered bases. |
| 2. 7. 1.- Phosphotransferases with an alcohol group as acceptor. | 3. 1.26.- Endoribonucleases producing 5'-phosphomonoesters. |
| 2. 7. 2.- Phosphotransferases with a carboxyl group as acceptor. | 3. 1.27.- Endoribonucleases producing other than 5'-phosphomonoesters. |
| 2. 7. 3.- Phosphotransferases with a nitrogenous group as acceptor. | 3. 1.30.- Endoribonucleases active with either ribo- or deoxyribonucleic and producing 5'-phosphomonoesters |
| 2. 7. 4.- Phosphotransferases with a phosphate group as acceptor. | 3. 1.31.- Endoribonucleases active with either ribo- or deoxyribonucleic and producing 3'-phosphomonoesters |
| 2. 7. 6.- Diphosphotransferases. | 3. 2. -.- Glycosylases. |
| 2. 7. 7.- Nucleotidyltransferases. | 3. 2. 1.- Glycosidases, i.e. enzymes hydrolyzing O- and S-glycosyl compounds |
| 2. 7. 8.- Transferases for other substituted phosphate groups. | 3. 2. 2.- Hydrolyzing N-glycosyl compounds. |
| 2. 7. 9.- Phosphotransferases with paired acceptors. | 3. 3. -.- Acting on ether bonds. |
| 2. 8. -.- Transferring sulfur-containing groups. | 3. 3. 1.- Thioether and trialkylsulfonium hydrolases. |
| 2. 8. 1.- Sulfurtransferases. | 3. 3. 2.- Ether hydrolases. |
| 2. 8. 2.- Sulfotransferases. | 3. 4. -.- Acting on peptide bonds (peptide hydrolases). |
| 2. 8. 3.- CoA-transferases. | 3. 4.11.- Aminopeptidases. |
| 2. 8. 4.- Transferring alkylthio groups. | 3. 4.13.- Dipeptidases. |

| | |
|---|---|
| 2. 9. -.- Transferring selenium-containing groups. | 3. 4.14.- Dipeptidyl-peptidases and tripeptidyl-peptidases. |
| 2. 9. 1.- Selenotransferases. | 3. 4.15.- Peptidyl-dipeptidases. |
| 3. -. -.- Hydrolases. | 3. 4.16.- Serine-type carboxypeptidases. |
| 3. 1. -.- Acting on ester bonds. | 3. 4.17.- Metallocarboxypeptidases. |
| 3. 1. 1.- Carboxylic ester hydrolases. | 3. 4.18.- Cysteine-type carboxypeptidases. |
| 3. 1. 2.- Thiolester hydrolases. | 3. 4.19.- Omega peptidases. |
| 3. 1. 3.- Phosphoric monoester hydrolases. | 3. 4.21.- Serine endopeptidases. |
| 3. 1. 4.- Phosphoric diester hydrolases. | 3. 4.22.- Cysteine endopeptidases. |
| 3. 1. 5.- Triphosphoric monoester hydrolases. | 3. 4.23.- Aspartic endopeptidases. |
| 3. 1. 6.- Sulfuric ester hydrolases. | 3. 4.24.- Metalloendopeptidases. |
| 3. 1. 7.- Diphosphoric monoester hydrolases. | 3. 4.25.- Threonine endopeptidases. |
| 3. 1. 8.- Phosphoric triester hydrolases. | 3. 4.99.- Endopeptidases of unknown catalytic mechanism. |
| 3. 1.11.- Exodeoxyribonucleases producing 5'-phosphomonoesters. | 3. 5. -.- Acting on carbon-nitrogen bonds, other than peptide bonds. |
| 3. 1.13.- Exoribonucleases producing 5'-phosphomonoesters. | 3. 5. 1.- In linear amides. |
| 3. 1.14.- Exoribonucleases producing 3'-phosphomonoesters. | 3. 5. 2.- In cyclic amides. |
| 3. 1.15.- Exonucleases active with either ribo- or deoxyribonucleic acid and producing 5'-phosphomonoesters | 3. 5. 3.- In linear amidines. |
| 3. 1.16.- Exonucleases active with either ribo- or deoxyribonucleic acid producing 3'-phosphomonoesters | 3. 5. 4.- In cyclic amidines. |
| 3. 1.21.- Endodeoxyribonucleases producing 5'-phosphomonoesters. | 3. 5. 5.- In nitriles. |

| | |
|---|---|
| 3. 5.99.- In other compounds. | 5. 3. -.- Intramolecular oxidoreductases. |
| 3. 6. -.- Acting on acid anhydrides. | 5. 3. 1.- Interconverting aldoses and ketoses. |
| 3. 6. 1.- In phosphorous-containing anhydrides. | 5. 3. 2.- Interconverting keto- and enol- groups. |
| 3. 6. 2.- In sulfonyl-containing anhydrides. | 5. 3. 3.- Transposing C==C bonds. |
| 3. 6. 3.- Acting on acid anhydrides; catalyzing transmembrane movement of substances | 5. 3. 4.- Transposing S-S bonds. |
| 3. 6. 4.- Acting on acid anhydrides; involved in cellular and subcellular movement | 5. 3.99.- Other intramolecular oxidoreductases. |
| 3. 6. 5.- Acting on GTP; involved in cellular and subcellular movement. | 5. 4. -.- Intramolecular transferases (mutases). |
| 3. 7. -.- Acting on carbon-carbon bonds. | 5. 4. 1.- Transferring acyl groups. |
| 3. 7. 1.- In ketonic substances. | 5. 4. 2.- Phosphotransferases (phosphomutases). |
| 3. 8. -.- Acting on halide bonds. | 5. 4. 3.- Transferring amino groups. |
| 3. 8. 1.- In C-halide compounds. | 5. 4. 4.- Transferring hydroxy groups. |
| 3. 9. -.- Acting on phosphorus-nitrogen bonds. | 5. 4.99.- Transferring other groups. |
| 3.10. -.- Acting on sulfur-nitrogen bonds. | 5. 5. -.- Intramolecular lyases. |
| 3.11. -.- Acting on carbon-phosphorus bonds. | 5.99. -.- Other isomerases. |
| 3.12. -.- Acting on sulfur-sulfur bonds. | 6. -. -.- Ligases. |
| 3.13. -.- Acting on carbon-sulfur bonds. | 6. 1. -.- Forming carbon-oxygen bonds. |
| 4. -. -.- Lyases. | 6. 1. 1.- Ligases forming aminoacyl-tRNA and related compounds. |
| 4. 1. -.- Carbon-carbon lyases. | 6. 2. -.- Forming carbon-sulfur bonds. |
| 4. 1. 1.- Carboxy-lyases. | 6. 2. 1.- Acid--thiol ligases. |
| 4. 1. 2.- Aldehyde-lyases. | 6. 3. -.- Forming carbon-nitrogen bonds. |
| 4. 1. 3.- Oxo-acid-lyases. | 6. 3. 1.- Acid--ammonia (or amide) ligases (amide synthases). |

| | |
|---|---|
| 4. 1.99.- Other carbon-carbon lyases. | 6. 3. 2.- Acid--D-amino-acid ligases (peptide synthases). |
| 4. 2. -.- Carbon-oxygen lyases. | 6. 3. 3.- Cyclo-ligases. |
| 4. 2. 1.- Hydro-lyases. | 6. 3. 4.- Other carbon--nitrogen ligases. |
| 4. 2. 2.- Acting on polysaccharides. | 6. 3. 5.- Carbon--nitrogen ligases with glutamine as amido-N-donor. |
| 4. 2. 3.- Acting on phosphates. | 6. 4. -.- Forming carbon-carbon bonds. |
| 4. 2.99.- Other carbon-oxygen lyases. | 6. 5. -.- Forming phosphoric ester bonds. |
| 4. 3. -.- Carbon-nitrogen lyases. | 6. 6. -.- Forming nitrogen-metal bonds. |
| 4. 3. 1.- Ammonia-lyases. | 6. 6. 1.- Forming nitrogen-metal bonds. |
| 4. 3. 2.- Lyases acting on amides, amidines, etc. | |
| 4. 3. 3.- Amine-lyases. | |
| 4. 3.99.- Other carbon-nitrogen-lyases. | |
| 4. 4. -.- Carbon-sulfur lyases. | |
| 4. 5. -.- Carbon-halide lyases. | |
| 4. 6. -.- Phosphorus-oxygen lyases. | |
| 4.99. -.- Other lyases. | |
| 5. -. -.- Isomerases. | |
| 5. 1. -.- Racemases and epimerases. | |
| 5. 1. 1.- Acting on amino acids and derivatives. | |
| 5. 1. 2.- Acting on hydroxy acids and derivatives. | |
| 5. 1. 3.- Acting on carbohydrates and derivatives. | |
| 5. 1.99.- Acting on other compounds. | |
| 5. 2. -.- Cis-trans-isomerases. | |

Table 3 (continued): EC Numbers with the corresponding name given to each enzyme class, subclass and sub-subclass.

| ENZYME: 1.-.-.- | 1.1.1.41 Isocitrate dehydrogenase (NAD+). |
|---|---|
| 1.1.1.1 Alcohol dehydrogenase. | 1.1.1.42 Isocitrate dehydrogenase (NADP+). |
| 1.1.1.2 Alcohol dehydrogenase (NADP+). | 1.1.1.43 Phosphogluconate 2-dehydrogenase. |
| 1.1.1.3 Homoserine dehydrogenase. | 1.1.1.44 Phosphogluconate dehydrogenase (decarboxylating). |
| 1.1.1.4 (R,R)-butanediol dehydrogenase. | 1.1.1.45 L-gulonate 3-dehydrogenase. |
| 1.1.1.5 Acetoin dehydrogenase. | 1.1.1.46 L-arabinose 1-dehydrogenase. |
| 1.1.1.6 Glycerol dehydrogenase. | 1.1.1.47 Glucose 1-dehydrogenase. |
| 1.1.1.7 Propanediol-phosphate dehydrogenase. | 1.1.1.48 Galactose 1-dehydrogenase. |
| 1.1.1.8 Glycerol-3-phosphate dehydrogenase (NAD+). | 1.1.1.49 Glucose-6-phosphate 1-dehydrogenase. |
| 1.1.1.9 D-xylulose reductase. | 1.1.1.50 3-alpha-hydroxysteroid dehydrogenase (B-specific). |
| 1.1.1.10 L-xylulose reductase. | 1.1.1.51 3(or 17)beta-hydroxysteroid dehydrogenase. |
| 1.1.1.11 D-arabinitol 4-dehydrogenase. | 1.1.1.52 3-alpha-hydroxycholanate dehydrogenase. |
| 1.1.1.12 L-arabinitol 4-dehydrogenase. | 1.1.1.53 3-alpha(or 20-beta)-hydroxysteroid dehydrogenase. |
| 1.1.1.13 L-arabinitol 2-dehydrogenase. | 1.1.1.54 Allyl-alcohol dehydrogenase. |
| 1.1.1.14 L-iditol 2-dehydrogenase. | 1.1.1.55 L-acetaldehyde reductase (NADPH). |
| 1.1.1.15 D-iditol 2-dehydrogenase. | 1.1.1.56 Ribitol 2-dehydrogenase. |
| 1.1.1.16 Galactitol 2-dehydrogenase. | 1.1.1.57 Fructuronate reductase. |
| 1.1.1.17 Mannitol-1-phosphate 5-dehydrogenase. | 1.1.1.58 Tagaturonate reductase. |
| 1.1.1.18 Inositol 2-dehydrogenase. | 1.1.1.59 3-hydroxypropionate dehydrogenase. |
| 1.1.1.19 L-glucuronate reductase. | 1.1.1.60 2-hydroxy-3-oxopropionate reductase. |

| | |
|---|---|
| 1.1.1.20 Glucuronolactone reductase. | 1.1.1.61 4-hydroxybutyrate dehydrogenase. |
| 1.1.1.21 Aldehyde reductase. | 1.1.1.62 Estradiol 17-beta-dehydrogenase. |
| 1.1.1.22 UDP-glucose 6-dehydrogenase. | 1.1.1.63 Testosterone 17-beta-dehydrogenase. |
| 1.1.1.23 Histidinol dehydrogenase. | 1.1.1.64 Testosterone 17-beta-dehydrogenase (NADP+). |
| 1.1.1.24 Quinate dehydrogenase. | 1.1.1.65 Pyridoxine 4-dehydrogenase. |
| 1.1.1.25 Shikimate dehydrogenase. | 1.1.1.66 Omega-hydroxydecanoate dehydrogenase. |
| 1.1.1.26 Glyoxylate reductase. | 1.1.1.67 Mannitol 2-dehydrogenase. |
| 1.1.1.27 L-lactate dehydrogenase. | 1.1.1.69 Gluconate 5-dehydrogenase. |
| 1.1.1.28 D-lactate dehydrogenase. | 1.1.1.71 Alcohol dehydrogenase (NAD(P)+). |
| 1.1.1.29 Glycerate dehydrogenase. | 1.1.1.72 Glycerol dehydrogenase (NADP+). |
| 1.1.1.30 3-hydroxybutyrate dehydrogenase. | 1.1.1.73 Octanol dehydrogenase. |
| 1.1.1.31 3-hydroxyisobutyrate dehydrogenase. | 1.1.1.75 (R)-aminopropanol dehydrogenase. |
| 1.1.1.32 Mevaldate reductase. | 1.1.1.76 (S,S)-butanediol dehydrogenase. |
| 1.1.1.33 Mevaldate reductase (NADPH). | 1.1.1.77 Lactaldehyde reductase. |
| 1.1.1.34 Hydroxymethylglutaryl-CoA reductase (NADPH). | 1.1.1.78 D-lactaldehyde dehydrogenase. |
| 1.1.1.35 3-hydroxyacyl-CoA dehydrogenase. | 1.1.1.79 Glyoxylate reductase (NADP+). |
| 1.1.1.36 Acetoacetyl-CoA reductase. | 1.1.1.80 Isopropanol dehydrogenase (NADP+). |
| 1.1.1.37 Malate dehydrogenase. | 1.1.1.81 Hydroxypyruvate reductase. |
| 1.1.1.38 Malate dehydrogenase (oxaloacetate-decarboxylating). | 1.1.1.82 Malate dehydrogenase (NADP+). |
| 1.1.1.39 Malate dehydrogenase (decarboxylating). | 1.1.1.83 D-malate dehydrogenase (decarboxylating). |
| 1.1.1.40 Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+). | 1.1.1.84 Dimethylmalate dehydrogenase. |

| | |
|---|---|
| 1.1.1.85 3-isopropylmalate dehydrogenase. | 1.1.1.127 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase. |
| 1.1.1.86 Ketol-acid reductoisomerase. | 1.1.1.128 L-idonate 2-dehydrogenase. |
| 1.1.1.87 Homoisocitrate dehydrogenase. | 1.1.1.129 L-threonate 3-dehydrogenase. |
| 1.1.1.88 Hydroxymethylglutaryl-CoA reductase. | 1.1.1.130 3-dehydro-L-gulonate 2-dehydrogenase. |
| 1.1.1.90 Aryl-alcohol dehydrogenase. | 1.1.1.131 Mannuronate reductase. |
| 1.1.1.91 Aryl-alcohol dehydrogenase (NADP+). | 1.1.1.132 GDP-mannose 6-dehydrogenase. |
| 1.1.1.92 Oxaloglycolate reductase (decarboxylating). | 1.1.1.133 dTDP-4-dehydrorhamnose reductase. |
| 1.1.1.93 Tartrate dehydrogenase. | 1.1.1.134 dTDP-6-deoxy-L-talose 4-dehydrogenase. |
| 1.1.1.94 Glycerol-3-phosphate dehydrogenase (NAD(P)+). | 1.1.1.135 GDP-6-deoxy-D-talose 4-dehydrogenase. |
| 1.1.1.95 Phosphoglycerate dehydrogenase. | 1.1.1.136 UDP-N-acetylglucosamine 6-dehydrogenase. |
| 1.1.1.96 Diiodophenylpyruvate reductase. | 1.1.1.137 Ribitol-5-phosphate 2-dehydrogenase. |
| 1.1.1.97 3-hydroxybenzyl-alcohol dehydrogenase. | 1.1.1.138 Mannitol 2-dehydrogenase (NADP+). |
| 1.1.1.98 (R)-2-hydroxy-fatty-acid dehydrogenase. | 1.1.1.140 Sorbitol-6-phosphate 2-dehydrogenase. |
| 1.1.1.99 (S)-2-hydroxy-fatty-acid dehydrogenase. | 1.1.1.141 15-hydroxyprostaglandin dehydrogenase (NAD+). |
| 1.1.1.100 3-oxoacyl-[acyl-carrier-protein] reductase. | 1.1.1.142 D-pinitol dehydrogenase. |
| 1.1.1.101 Acylglycerone-phosphate reductase. | 1.1.1.143 Sequoyitol dehydrogenase. |
| 1.1.1.102 3-dehydrosphinganine reductase. | 1.1.1.144 Perillyl-alcohol dehydrogenase. |
| 1.1.1.103 L-threonine 3-dehydrogenase. | 1.1.1.145 3-beta-hydroxy-delta(5)-steroid dehydrogenase. |

| | |
|---|---|
| 1.1.1.104 4-oxoproline reductase. | 1.1.1.146 11-beta-hydroxysteroid dehydrogenase. |
| 1.1.1.105 Retinol dehydrogenase. | 1.1.1.147 16-alpha-hydroxysteroid dehydrogenase. |
| 1.1.1.106 Pantoate 4-dehydrogenase. | 1.1.1.148 Estradiol 17-alpha-dehydrogenase. |
| 1.1.1.107 Pyridoxal 4-dehydrogenase. | 1.1.1.149 20-alpha-hydroxysteroid dehydrogenase. |
| 1.1.1.108 Carnitine 3-dehydrogenase. | 1.1.1.150 21-hydroxysteroid dehydrogenase (NAD+). |
| 1.1.1.110 Indolelactate dehydrogenase. | 1.1.1.152 3-alpha-hydroxy-5-beta-androstane-17-one 3-alpha-dehydrogenase. |
| 1.1.1.111 3-(imidazol-5-yl)lactate dehydrogenase. | 1.1.1.153 Sepiapterin reductase. |
| 1.1.1.112 Indanol dehydrogenase. | 1.1.1.154 Ureidoglycolate dehydrogenase. |
| 1.1.1.113 L-xylose 1-dehydrogenase. | 1.1.1.155 Homoisocitrate dehydrogenase. |
| 1.1.1.114 Apiose 1-reductase. | 1.1.1.156 Glycerol 2-dehydrogenase (NADP+). |
| 1.1.1.115 Ribose 1-dehydrogenase (NADP+). | 1.1.1.157 3-hydroxybutyryl-CoA dehydrogenase. |
| 1.1.1.116 D-arabinose 1-dehydrogenase. | 1.1.1.158 UDP-N-acetylmuramate dehydrogenase. |
| 1.1.1.117 D-arabinose 1-dehydrogenase (NAD(P)+). | 1.1.1.159 7-alpha-hydroxysteroid dehydrogenase. |
| 1.1.1.118 Glucose 1-dehydrogenase (NAD+). | 1.1.1.160 Dihydrobunolol dehydrogenase. |
| 1.1.1.119 Glucose 1-dehydrogenase (NADP+). | 1.1.1.161 Cholestanetetraol 26-dehydrogenase. |
| 1.1.1.120 Galactose 1-dehydrogenase (NADP+). | 1.1.1.162 Erythrulose reductase. |
| 1.1.1.121 Aldose 1-dehydrogenase. | 1.1.1.163 Cyclopentanol dehydrogenase. |
| 1.1.1.122 D-threo-aldose 1-dehydrogenase. | 1.1.1.164 Hexadecanol dehydrogenase. |
| 1.1.1.123 Sorbose 5-dehydrogenase (NADP+). | 1.1.1.165 2-alkyn-1-ol dehydrogenase. |
| 1.1.1.124 Fructose 5-dehydrogenase (NADP+). | 1.1.1.166 Hydroxycyclohexanecarboxylate dehydrogenase. |
| 1.1.1.125 2-deoxy-D-gluconate 3-dehydrogenase. | 1.1.1.167 Hydroxymalonate dehydrogenase. |
| 1.1.1.126 2-dehydro-3-deoxy-D-gluconate 6-dehydrogenase. | 1.1.1.168 2-dehydropantolactone reductase (A-specific). |

| | |
|---|---|
| 1.1.1.169 2-dehydropantoate 2-reductase. | 1.1.1.207 (-)-menthol dehydrogenase. |
| 1.1.1.170 Sterol-4-alpha-carboxylate 3-dehydrogenase (decarboxylating). | 1.1.1.208 (+)-neomenthol dehydrogenase. |
| 1.1.1.172 2-oxoadipate reductase. | 1.1.1.209 3(or 17)-alpha-hydroxysteroid dehydrogenase. |
| 1.1.1.173 L-rhamnose 1-dehydrogenase. | 1.1.1.210 3-beta(or 20-alpha)-hydroxysteroid dehydrogenase. |
| 1.1.1.174 Cyclohexane-1,2-diol dehydrogenase. | 1.1.1.211 Long-chain-3-hydroxyacyl-CoA dehydrogenase. |
| 1.1.1.175 D-xylose 1-dehydrogenase. | 1.1.1.212 3-oxoacyl-[acyl-carrier-protein] reductase (NADH). |
| 1.1.1.176 12-alpha-hydroxysteroid dehydrogenase. | 1.1.1.213 3-alpha-hydroxysteroid dehydrogenase (A-specific). |
| 1.1.1.177 Glycerol-3-phosphate 1-dehydrogenase (NADP+). | 1.1.1.214 2-dehydropantolactone reductase (B-specific). |
| 1.1.1.178 3-hydroxy-2-methylbutyryl-CoA dehydrogenase. | 1.1.1.215 Gluconate 2-dehydrogenase. |
| 1.1.1.179 D-xylose 1-dehydrogenase (NADP+). | 1.1.1.216 Farnesol dehydrogenase. |
| 1.1.1.181 Cholest-5-ene-3-beta,7-alpha-diol 3-beta-dehydrogenase. | 1.1.1.217 Benzyl-2-methyl-hydroxybutyrate dehydrogenase. |
| 1.1.1.183 Geraniol dehydrogenase. | 1.1.1.218 Morphine 6-dehydrogenase. |
| 1.1.1.184 Carbonyl reductase (NADPH). | 1.1.1.219 Dihydrokaempferol 4-reductase. |
| 1.1.1.185 L-glycol dehydrogenase. | 1.1.1.220 6-pyruvoyltetrahydropterin 2'-reductase. |
| 1.1.1.186 dTDP-galactose 6-dehydrogenase. | 1.1.1.221 Vomifoliol 4'-dehydrogenase. |
| 1.1.1.187 GDP-4-dehydro-D-rhamnose reductase. | 1.1.1.222 (R)-4-hydroxyphenyllactate dehydrogenase. |

| | |
|---|---|
| 1.1.1.188 Prostaglandin-F synthase. | 1.1.1.223 Isopiperitenol dehydrogenase. |
| 1.1.1.189 Prostaglandin-E(2) 9-reductase. | 1.1.1.224 Mannose-6-phosphate 6-reductase. |
| 1.1.1.190 Indole-3-acetaldehyde reductase (NADH). | 1.1.1.225 Chlordecone reductase. |
| 1.1.1.191 Indole-3-acetaldehyde reductase (NADPH). | 1.1.1.226 4-hydroxycyclohexanecarboxylate dehydrogenase. |
| 1.1.1.192 Long-chain-alcohol dehydrogenase. | 1.1.1.227 (-)-borneol dehydrogenase. |
| 1.1.1.193 5-amino-6-(5-phosphoribosylamino)uracil reductase. | 1.1.1.228 (+)-sabinol dehydrogenase. |
| 1.1.1.194 Coniferyl-alcohol dehydrogenase. | 1.1.1.229 Diethyl 2-methyl-3-oxosuccinate reductase. |
| 1.1.1.195 Cinnamyl-alcohol dehydrogenase. | 1.1.1.230 3-alpha-hydroxyglycyrrhetinate dehydrogenase. |
| 1.1.1.196 15-hydroxyprostaglandin-D dehydrogenase (NADP+). | 1.1.1.231 15-hydroxyprostaglandin-I dehydrogenase (NADP+). |
| 1.1.1.197 15-hydroxyprostaglandin dehydrogenase (NADP+). | 1.1.1.232 15-hydroxyicosatetraenoate dehydrogenase. |
| 1.1.1.198 (+)-borneol dehydrogenase. | 1.1.1.233 N-acylmannosamine 1-dehydrogenase. |
| 1.1.1.199 (S)-usnate reductase. | 1.1.1.234 Flavanone 4-reductase. |
| 1.1.1.200 Aldose-6-phosphate reductase (NADPH). | 1.1.1.235 8-oxocoformycin reductase. |
| 1.1.1.201 7-beta-hydroxysteroid dehydrogenase (NADP+). | 1.1.1.236 Tropinone reductase. |
| 1.1.1.202 1,3-propanediol dehydrogenase. | 1.1.1.237 Hydroxyphenylpyruvate reductase. |
| 1.1.1.203 Uronate dehydrogenase. | 1.1.1.238 12-beta-hydroxysteroid dehydrogenase. |
| 1.1.1.205 IMP dehydrogenase. | 1.1.1.239 3-alpha-(17-beta)-hydroxysteroid dehydrogenase (NAD+). |
| 1.1.1.206 Tropine dehydrogenase. | 1.1.1.240 N-acetylhexosamine 1-dehydrogenase. |

| | |
|---|---|
| 1.1.1.241 6-endo-hydroxycineole dehydrogenase. | 1.1.1.281 GDP-4-dehydro-6-deoxy-D-mannose reductase. |
| 1.1.1.243 Carveol dehydrogenase. | 1.1.1.282 Quinate/shikimate dehydrogenase. |
| 1.1.1.244 Methanol dehydrogenase. | 1.1.2.2 Mannitol dehydrogenase (cytochrome). |
| 1.1.1.245 Cyclohexanol dehydrogenase. | 1.1.2.3 L-lactate dehydrogenase (cytochrome). |
| 1.1.1.246 Pterocarpin synthase. | 1.1.2.4 D-lactate dehydrogenase (cytochrome). |
| 1.1.1.247 Codeinone reductase (NADPH). | 1.1.2.5 D-lactate dehydrogenase (cytochrome c-553). |
| 1.1.1.248 Salutaridine reductase (NADPH). | 1.1.3.3 Malate oxidase. |
| 1.1.1.250 D-arabinitol 2-dehydrogenase. | 1.1.3.4 Glucose oxidase. |
| 1.1.1.251 Galactitol-1-phosphate 5-dehydrogenase. | 1.1.3.5 Hexose oxidase. |
| 1.1.1.252 Tetrahydroxynaphthalene reductase. | 1.1.3.6 Cholesterol oxidase. |
| 1.1.1.254 (S)-carnitine 3-dehydrogenase. | 1.1.3.7 Aryl-alcohol oxidase. |
| 1.1.1.255 Mannitol dehydrogenase. | 1.1.3.8 L-gulonolactone oxidase. |
| 1.1.1.256 Fluoren-9-ol dehydrogenase. | 1.1.3.9 Galactose oxidase. |
| 1.1.1.257 4-(hydroxymethyl)benzenesulfonate dehydrogenase. | 1.1.3.10 Pyranose oxidase. |
| 1.1.1.258 6-hydroxyhexanoate dehydrogenase. | 1.1.3.11 L-sorbose oxidase. |
| 1.1.1.259 3-hydroxypimeloyl-CoA dehydrogenase. | 1.1.3.12 Pyridoxine 4-oxidase. |
| 1.1.1.260 Sulcatone reductase. | 1.1.3.13 Alcohol oxidase. |
| 1.1.1.261 Glycerol-1-phosphate dehydrogenase (NAD(P)+). | 1.1.3.14 Catechol oxidase (dimerizing). |

| | |
|---|---|
| 1.1.1.262 4-hydroxythreonine-4-phosphate dehydrogenase. | 1.1.3.15 (S)-2-hydroxy-acid oxidase. |
| 1.1.1.263 1,5-anhydro-D-fructose reductase. | 1.1.3.16 Ecdysone oxidase. |
| 1.1.1.264 L-idonate 5-dehydrogenase. | 1.1.3.17 Choline oxidase. |
| 1.1.1.265 3-methylbutanal reductase. | 1.1.3.18 Secondary-alcohol oxidase. |
| 1.1.1.266 dTDP-4-dehydro-6-deoxyglucose reductase. | 1.1.3.19 4-hydroxymandelate oxidase. |
| 1.1.1.267 1-deoxy-D-xylulose-5-phosphate reductoisomerase. | 1.1.3.20 Long-chain-alcohol oxidase. |
| 1.1.1.268 2-(R)-hydroxypropyl-CoM dehydrogenase. | 1.1.3.21 Glycerol-3-phosphate oxidase. |
| 1.1.1.269 2-(S)-hydroxypropyl-CoM dehydrogenase. | 1.1.3.23 Thiamine oxidase. |
| 1.1.1.270 3-keto-steroid reductase. | 1.1.3.24 L-galactonolactone oxidase. |
| 1.1.1.271 GDP-L-fucose synthase. | 1.1.3.25 Cellobiose oxidase. |
| 1.1.1.272 (R)-2-hydroxyacid dehydrogenase. | 1.1.3.27 Hydroxyphytanate oxidase. |
| 1.1.1.273 Vellosimine dehydrogenase. | 1.1.3.28 Nucleoside oxidase. |
| 1.1.1.274 2,5-didehydrogluconate reductase. | 1.1.3.29 N-acylhexosamine oxidase. |
| 1.1.1.275 (+)-trans-carveol dehydrogenase. | 1.1.3.30 Polyvinyl-alcohol oxidase. |
| 1.1.1.276 Serine 3-dehydrogenase. | 1.1.3.37 D-arabinono-1,4-lactone oxidase. |
| 1.1.1.277 3-beta-hydroxy-5-beta-steroid dehydrogenase. | 1.1.3.38 Vanillyl-alcohol oxidase. |
| 1.1.1.278 3-beta-hydroxy-5-alpha-steroid dehydrogenase. | 1.1.3.39 Nucleoside oxidase (H(2)O(2)-forming). |
| 1.1.1.279 (R)-3-hydroxyacid-ester dehydrogenase. | 1.1.3.40 D-mannitol oxidase. |
| 1.1.1.280 (S)-3-hydroxyacid-ester dehydrogenase. | 1.1.3.41 Xylitol oxidase. |

| | |
|---|---|
| 1.1.4.1 Vitamin-K-epoxide reductase (warfarin-sensitive). | 1.2.1.8 Betaine-aldehyde dehydrogenase. |
| 1.1.4.2 Vitamin-K-epoxide reductase (warfarin-insensitive). | 1.2.1.9 Glyceraldehyde-3-phosphate dehydrogenase (NADP+). |
| 1.1.5.2 Quinoprotein glucose dehydrogenase. | 1.2.1.10 Acetaldehyde dehydrogenase (acetylating). |
| 1.1.99.1 Choline dehydrogenase. | 1.2.1.11 Aspartate-semialdehyde dehydrogenase. |
| 1.1.99.2 2-hydroxyglutarate dehydrogenase. | 1.2.1.12 Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating). |
| 1.1.99.3 Gluconate 2-dehydrogenase (acceptor). | 1.2.1.13 Glyceraldehyde-3-phosphate dehydrogenase (NADP(+)) (phosphorylating). |
| 1.1.99.4 Dehydrogluconate dehydrogenase. | 1.2.1.15 Malonate-semialdehyde dehydrogenase. |
| 1.1.99.5 Glycerol-3-phosphate dehydrogenase. | 1.2.1.16 Succinate-semialdehyde dehydrogenase (NAD(P)+). |
| 1.1.99.6 D-2-hydroxy-acid dehydrogenase. | 1.2.1.17 Glyoxylate dehydrogenase (acylating). |
| 1.1.99.7 Lactate--malate transhydrogenase. | 1.2.1.18 Malonate-semialdehyde dehydrogenase (acetylating). |
| 1.1.99.8 Alcohol dehydrogenase (acceptor). | 1.2.1.19 Aminobutyraldehyde dehydrogenase. |
| 1.1.99.9 Pyridoxine 5-dehydrogenase. | 1.2.1.20 Glutarate-semialdehyde dehydrogenase. |
| 1.1.99.10 Glucose dehydrogenase (acceptor). | 1.2.1.21 Glycolaldehyde dehydrogenase. |
| 1.1.99.11 Fructose 5-dehydrogenase. | 1.2.1.22 Lactaldehyde dehydrogenase. |
| 1.1.99.12 Sorbose dehydrogenase. | 1.2.1.23 2-oxoaldehyde dehydrogenase (NAD+). |
| 1.1.99.13 Glucoside 3-dehydrogenase. | 1.2.1.24 Succinate-semialdehyde dehydrogenase. |
| 1.1.99.14 Glycolate dehydrogenase. | 1.2.1.25 2-oxoisovalerate dehydrogenase (acylating). |
| 1.1.99.16 Malate dehydrogenase (acceptor). | 1.2.1.26 2,5-dioxovalerate dehydrogenase. |
| 1.1.99.18 Cellobiose dehydrogenase (acceptor). | 1.2.1.27 Methylmalonate-semialdehyde dehydrogenase (acylating). |
| 1.1.99.19 Uracil dehydrogenase. | 1.2.1.28 Benzaldehyde dehydrogenase (NAD+). |

| | |
|---|---|
| 1.1.99.20 Alkan-1-ol dehydrogenase (acceptor). | 1.2.1.29 Aryl-aldehyde dehydrogenase. |
| 1.1.99.21 D-sorbitol dehydrogenase (acceptor). | 1.2.1.30 Aryl-aldehyde dehydrogenase (NADP+). |
| 1.1.99.22 Glycerol dehydrogenase (acceptor). | 1.2.1.31 L-aminoadipate-semialdehyde dehydrogenase. |
| 1.1.99.23 Polyvinyl-alcohol dehydrogenase (acceptor). | 1.2.1.32 Aminomuconate-semialdehyde dehydrogenase. |
| 1.1.99.24 Hydroxyacid--oxoacid transhydrogenase. | 1.2.1.33 (R)-dehydropantoate dehydrogenase. |
| 1.1.99.25 Quinate dehydrogenase (pyrroloquinoline-quinone). | 1.2.1.36 Retinal dehydrogenase. |
| 1.1.99.26 3-hydroxycyclohexanone dehydrogenase. | 1.2.1.38 N-acetyl-gamma-glutamyl-phosphate reductase. |
| 1.1.99.27 (R)-pantolactone dehydrogenase (flavin). | 1.2.1.39 Phenylacetaldehyde dehydrogenase. |
| 1.1.99.28 Glucose--fructose oxidoreductase. | 1.2.1.40 3-alpha,7-alpha,12-alpha-trihydroxycholestan-26-al 26-oxidoreductase. |
| 1.1.99.29 Pyranose dehydrogenase (acceptor). | 1.2.1.41 Glutamate-5-semialdehyde dehydrogenase. |
| 1.1.99.30 2-oxo-acid reductase. | 1.2.1.42 Hexadecanal dehydrogenase (acylating). |
| 1.2.1.1 Formaldehyde dehydrogenase (glutathione). | 1.2.1.43 Formate dehydrogenase (NADP+). |
| 1.2.1.2 Formate dehydrogenase. | 1.2.1.44 Cinnamoyl-CoA reductase. |
| 1.2.1.3 Aldehyde dehydrogenase (NAD+). | 1.2.1.45 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase. |
| 1.2.1.4 Aldehyde dehydrogenase (NADP+). | 1.2.1.46 Formaldehyde dehydrogenase. |
| 1.2.1.5 Aldehyde dehydrogenase (NAD(P)+). | 1.2.1.47 4-trimethylammoniobutyraldehyde dehydrogenase. |
| 1.2.1.7 Benzaldehyde dehydrogenase (NADP+). | 1.2.1.48 Long-chain-aldehyde dehydrogenase. |
| 1.2.1.49 2-oxoaldehyde dehydrogenase (NADP+). | 1.2.4.4 3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring). |
| 1.2.1.50 Long-chain-fatty-acyl-CoA reductase. | 1.2.7.1 Pyruvate synthase. |
| 1.2.1.51 Pyruvate dehydrogenase (NADP+). | 1.2.7.2 2-oxobutyrate synthase. |
| 1.2.1.52 Oxoglutarate dehydrogenase (NADP+). | 1.2.7.3 2-oxoglutarate synthase. |
| 1.2.1.53 4-hydroxyphenylacetaldehyde dehydrogenase. | 1.2.7.4 Carbon-monoxide dehydrogenase (ferredoxin). |
| 1.2.1.54 Gamma-guanidinobutyraldehyde dehydrogenase. | 1.2.7.5 Aldehyde ferredoxin oxidoreductase. |
| 1.2.1.57 Butanal dehydrogenase. | 1.2.7.6 Glyceraldehyde-3-phosphate dehydrogenase (ferredoxin). |
| 1.2.1.58 Phenylglyoxylate dehydrogenase (acylating). | 1.2.7.7 3-methyl-2-oxobutanoate dehydrogenase (ferredoxin). |
| 1.2.1.59 Glyceraldehyde-3-phosphate dehydrogenase (NAD(P)(+)) (phosphorylating). | 1.2.7.8 Indolepyruvate ferredoxin oxidoreductase. |
| 1.2.1.60 5-carboxymethyl-2-hydroxymuconic-semialdehyde dehydrogenase. | 1.2.7.9 2-oxoglutarate ferredoxin oxidoreductase. |
| 1.2.1.61 4-hydroxymuconic-semialdehyde dehydrogenase. | 1.2.99.2 Carbon-monoxide dehydrogenase (acceptor). |
| 1.2.1.62 4-formylbenzenesulfonate dehydrogenase. | 1.2.99.3 Aldehyde dehydrogenase (pyrroloquinoline-quinone). |
| 1.2.1.63 6-oxohexanoate dehydrogenase. | 1.2.99.4 Formaldehyde dismutase. |
| 1.2.1.64 4-hydroxybenzaldehyde dehydrogenase. | 1.2.99.5 Formylmethanofuran dehydrogenase. |
| 1.2.1.65 Salicylaldehyde dehydrogenase. | 1.2.99.6 Carboxylate reductase. |
| 1.2.1.66 Mycothiol-dependent formaldehyde dehydrogenase. | 1.2.99.7 Aldehyde dehydrogenase (FAD-independent). |
| 1.2.1.67 Vanillin dehydrogenase. | 1.3.1.1 Dihydrouracil dehydrogenase (NAD+). |

| | |
|---|---|
| 1.2.1.68 Coniferyl-aldehyde dehydrogenase. | 1.3.1.2 Dihydropyrimidine dehydrogenase (NADP+). |
| 1.2.1.69 Fluoroacetaldehyde dehydrogenase. | 1.3.1.3 Cortisone beta-reductase. |
| 1.2.2.1 Formate dehydrogenase (cytochrome). | 1.3.1.4 Cortisone alpha-reductase. |
| 1.2.2.2 Pyruvate dehydrogenase (cytochrome). | 1.3.1.5 Cucurbitacin delta(23)-reductase. |
| 1.2.2.3 Formate dehydrogenase (cytochrome c-553). | 1.3.1.6 Fumarate reductase (NADH). |
| 1.2.2.4 Carbon-monoxide dehydrogenase (cytochrome b-561). | 1.3.1.7 Meso-tartrate dehydrogenase. |
| 1.2.3.1 Aldehyde oxidase. | 1.3.1.8 Acyl-CoA dehydrogenase (NADP+). |
| 1.2.3.3 Pyruvate oxidase. | 1.3.1.9 Enoyl-[acyl-carrier-protein] reductase (NADH). |
| 1.2.3.4 Oxalate oxidase. | 1.3.1.10 Enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific). |
| 1.2.3.5 Glyoxylate oxidase. | 1.3.1.11 2-coumarate reductase. |
| 1.2.3.6 Pyruvate oxidase (CoA-acetylating). | 1.3.1.12 Prephenate dehydrogenase. |
| 1.2.3.7 Indole-3-acetaldehyde oxidase. | 1.3.1.13 Prephenate dehydrogenase (NADP+). |
| 1.2.3.8 Pyridoxal oxidase. | 1.3.1.14 Orotate reductase (NADH). |
| 1.2.3.9 Aryl-aldehyde oxidase. | 1.3.1.15 Orotate reductase (NADPH). |
| 1.2.3.11 Retinal oxidase. | 1.3.1.16 Beta-nitroacrylate reductase. |
| 1.2.3.13 4-hydroxyphenylpyruvate oxidase. | 1.3.1.17 3-methyleneoxindole reductase. |
| 1.2.4.1 Pyruvate dehydrogenase (acetyl-transferring). | 1.3.1.18 Kynurenate-7,8-dihydrodiol dehydrogenase. |
| 1.2.4.2 Oxoglutarate dehydrogenase (succinyl-transferring). | 1.3.1.19 Cis-1,2-dihydrobenzene-1,2-diol dehydrogenase. |

| | |
|---|---|
| 1.3.1.20 Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase. | 1.3.1.59 1,6-dihydroxy-5-methylcyclohexa-2,4-dienecarboxylate dehydrogenase. |
| 1.3.1.21 7-dehydrocholesterol reductase. | 1.3.1.60 Dibenzothiophene dihydrodiol dehydrogenase. |
| 1.3.1.22 Cholestenone 5-alpha-reductase. | 1.3.1.61 Terephthalate 1,2-cis-dihydrodiol dehydrogenase. |
| 1.3.1.23 Cholestenone 5-beta-reductase. | 1.3.1.62 Pimeloyl-CoA dehydrogenase. |
| 1.3.1.24 Biliverdin reductase. | 1.3.1.63 2,4-dichlorobenzoyl-CoA reductase. |
| 1.3.1.25 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase. | 1.3.1.64 Phthalate 4,5-cis-dihydrodiol dehydrogenase. |
| 1.3.1.26 Dihydrodipicolinate reductase. | 1.3.1.65 5,6-dihydroxy-3-methyl-2-oxo-1,2,5,6-tetrahydroquinoline dehydrogenase. |
| 1.3.1.27 2-hexadecenal reductase. | 1.3.1.66 Cis-dihydroethylcatechol dehydrogenase. |
| 1.3.1.28 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. | 1.3.1.67 Cis-1,2-dihydroxy-4-methylcyclohexa-3,5-diene-1-carboxylate dehydrogenase. |
| 1.3.1.29 Cis-1,2-dihydro-1,2-dihydroxynaphthalene dehydrogenase. | 1.3.1.68 1,2-dihydroxy-6-methylcyclohexa-3,5-dienecarboxylate dehydrogenase. |
| 1.3.1.30 Progesterone 5-alpha-reductase. | 1.3.1.69 Zeatin reductase. |
| 1.3.1.31 2-enoate reductase. | 1.3.1.70 Delta(14)-sterol reductase. |
| 1.3.1.32 Maleylacetate reductase. | 1.3.1.71 Delta(24(24(1)))-sterol reductase. |
| 1.3.1.33 Protochlorophyllide reductase. | 1.3.1.72 Delta(24)-sterol reductase. |
| 1.3.1.34 2,4-dienoyl-CoA reductase (NADPH). | 1.3.1.73 1,2-dihydrovomilenine reductase. |
| 1.3.1.35 Phosphatidylcholine desaturase. | 1.3.1.74 2-alkenal reductase. |
| 1.3.1.36 Geissoschizine dehydrogenase. | 1.3.1.75 Divinyl chlorophyllide a 8-vinyl-reductase. |
| 1.3.1.37 Cis-2-enoyl-CoA reductase (NADPH). | 1.3.1.76 Precorrin-2 dehydrogenase. |
| 1.3.1.38 Trans-2-enoyl-CoA reductase (NADPH). | 1.3.2.3 Galactonolactone dehydrogenase. |

| | |
|---|---|
| 1.3.1.39 Enoyl-[acyl-carrier-protein] reductase (NADPH, A-specific). | 1.3.3.1 Dihydroorotate oxidase. |
| 1.3.1.40 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate reductase. | 1.3.3.2 Lathosterol oxidase. |
| 1.3.1.41 Xanthommatin reductase. | 1.3.3.3 Coproporphyrinogen oxidase. |
| 1.3.1.42 12-oxophytodienoate reductase. | 1.3.3.4 Protoporphyrinogen oxidase. |
| 1.3.1.43 Cyclohexadienyl dehydrogenase. | 1.3.3.5 Bilirubin oxidase. |
| 1.3.1.44 Trans-2-enoyl-CoA reductase (NAD+). | 1.3.3.6 Acyl-CoA oxidase. |
| 1.3.1.45 2'-hydroxyisoflavone reductase. | 1.3.3.7 Dihydrouracil oxidase. |
| 1.3.1.46 Biochanin-A reductase. | 1.3.3.8 Tetrahydroberberine oxidase. |
| 1.3.1.47 Alpha-santonin 1,2-reductase. | 1.3.3.9 Secologanin synthase. |
| 1.3.1.48 15-oxoprostaglandin 13-oxidase. | 1.3.3.10 Tryptophan alpha,beta-oxidase. |
| 1.3.1.49 Cis-3,4-dihydrophenanthrene-3,4-diol dehydrogenase. | 1.3.5.1 Succinate dehydrogenase (ubiquinone). |
| 1.3.1.51 2'-hydroxydaidzein reductase. | 1.3.7.1 6-hydroxynicotinate reductase. |
| 1.3.1.52 2-methyl-branched-chain-enoyl-CoA reductase. | 1.3.7.2 15,16-dihydrobiliverdin:ferredoxin oxidoreductase. |
| 1.3.1.53 (3S,4R)-3,4-dihydroxycyclohexa-1,5-diene-1,4-dicarboxylate dehydrogenase. | 1.3.7.3 Phycoerythrobilin:ferredoxin oxidoreductase. |
| 1.3.1.54 Precorrin-6A reductase. | 1.3.7.4 Phytochromobilin:ferredoxin oxidoreductase. |
| 1.3.1.56 Cis-2,3-dihydrobiphenyl-2,3-diol dehydrogenase. | 1.3.7.5 Phycocyanobilin:ferredoxin oxidoreductase. |
| 1.3.1.57 Phloroglucinol reductase. | 1.3.99.1 Succinate dehydrogenase. |
| 1.3.1.58 2,3-dihydroxy-2,3-dihydro-p-cumate dehydrogenase. | 1.3.99.2 Butyryl-CoA dehydrogenase. |

| | |
|---|---|
| 1.3.99.3 Acyl-CoA dehydrogenase. | 1.4.3.5 Pyridoxamine-phosphate oxidase. |
| 1.3.99.4 3-oxosteroid 1-dehydrogenase. | 1.4.3.6 Amine oxidase (copper-containing). |
| 1.3.99.5 3-oxo-5-alpha-steroid 4-dehydrogenase. | 1.4.3.7 D-glutamate oxidase. |
| 1.3.99.6 3-oxo-5-beta-steroid 4-dehydrogenase. | 1.4.3.8 Ethanolamine oxidase. |
| 1.3.99.7 Glutaryl-CoA dehydrogenase. | 1.4.3.10 Putrescine oxidase. |
| 1.3.99.8 2-furoyl-CoA dehydrogenase. | 1.4.3.11 L-glutamate oxidase. |
| 1.3.99.10 Isovaleryl-CoA dehydrogenase. | 1.4.3.12 Cyclohexylamine oxidase. |
| 1.3.99.11 Dihydroorotate dehydrogenase. | 1.4.3.13 Protein-lysine 6-oxidase. |
| 1.3.99.12 2-methylacyl-CoA dehydrogenase. | 1.4.3.14 L-lysine oxidase. |
| 1.3.99.13 Long-chain-acyl-CoA dehydrogenase. | 1.4.3.15 D-glutamate(D-aspartate) oxidase. |
| 1.3.99.14 Cyclohexanone dehydrogenase. | 1.4.3.16 L-aspartate oxidase. |
| 1.3.99.15 Benzoyl-CoA reductase. | 1.4.3.19 Glycine oxidase. |
| 1.3.99.16 Isoquinoline 1-oxidoreductase. | 1.4.4.2 Glycine dehydrogenase (decarboxylating). |
| 1.3.99.17 Quinoline 2-oxidoreductase. | 1.4.7.1 Glutamate synthase (ferredoxin). |
| 1.3.99.18 Quinaldate 4-oxidoreductase. | 1.4.99.1 D-amino-acid dehydrogenase. |
| 1.3.99.19 Quinoline-4-carboxylate 2-oxidoreductase. | 1.4.99.2 Taurine dehydrogenase. |
| 1.3.99.20 4-hydroxybenzoyl-CoA reductase. | 1.4.99.3 Amine dehydrogenase. |
| 1.3.99.21 (R)-benzylsuccinyl-CoA dehydrogenase. | 1.4.99.4 Aralkylamine dehydrogenase. |
| 1.4.1.1 Alanine dehydrogenase. | 1.4.99.5 Glycine dehydrogenase (cyanide-forming). |
| 1.4.1.2 Glutamate dehydrogenase. | 1.5.1.1 Pyrroline-2-carboxylate reductase. |
| 1.4.1.3 Glutamate dehydrogenase (NAD(P)+). | 1.5.1.2 Pyrroline-5-carboxylate reductase. |
| 1.4.1.4 Glutamate dehydrogenase (NADP+). | 1.5.1.3 Dihydrofolate reductase. |
| 1.4.1.5 L-amino-acid dehydrogenase. | 1.5.1.5 Methylenetetrahydrofolate dehydrogenase (NADP+). |

| | |
|---|---|
| 1.4.1.7 Serine 2-dehydrogenase. | 1.5.1.6 Formyltetrahydrofolate dehydrogenase. |
| 1.4.1.8 Valine dehydrogenase (NADP+). | 1.5.1.7 Saccharopine dehydrogenase (NAD+, L-lysine-forming). |
| 1.4.1.9 Leucine dehydrogenase. | 1.5.1.8 Saccharopine dehydrogenase (NADP+, L-lysine-forming). |
| 1.4.1.10 Glycine dehydrogenase. | 1.5.1.9 Saccharopine dehydrogenase (NAD+, L-glutamate-forming). |
| 1.4.1.11 L-erythro-3,5-diaminohexanoate dehydrogenase. | 1.5.1.10 Saccharopine dehydrogenase (NADP+, L-glutamate-forming). |
| 1.4.1.12 2,4-diaminopentanoate dehydrogenase. | 1.5.1.11 D-octopine dehydrogenase. |
| 1.4.1.13 Glutamate synthase (NADPH). | 1.5.1.12 1-pyrroline-5-carboxylate dehydrogenase. |
| 1.4.1.14 Glutamate synthase (NADH). | 1.5.1.15 Methylenetetrahydrofolate dehydrogenase (NAD+). |
| 1.4.1.15 Lysine dehydrogenase. | 1.5.1.16 D-lysopine dehydrogenase. |
| 1.4.1.16 Diaminopimelate dehydrogenase. | 1.5.1.17 Alanopine dehydrogenase. |
| 1.4.1.17 N-methylalanine dehydrogenase. | 1.5.1.18 Ephedrine dehydrogenase. |
| 1.4.1.18 Lysine 6-dehydrogenase. | 1.5.1.19 D-nopaline dehydrogenase. |
| 1.4.1.19 Tryptophan dehydrogenase. | 1.5.1.20 Methylenetetrahydrofolate reductase (NADPH). |
| 1.4.1.20 Phenylalanine dehydrogenase. | 1.5.1.21 Delta(1)-piperideine-2-carboxylate reductase. |
| 1.4.2.1 Glycine dehydrogenase (cytochrome). | 1.5.1.22 Strombine dehydrogenase. |
| 1.4.3.1 D-aspartate oxidase. | 1.5.1.23 Tauropine dehydrogenase. |
| 1.4.3.2 L-amino-acid oxidase. | 1.5.1.24 N(5)-(carboxyethyl)ornithine synthase. |
| 1.4.3.3 D-amino-acid oxidase. | 1.5.1.25 Thiomorpholine-carboxylate dehydrogenase. |
| 1.4.3.4 Amine oxidase (flavin-containing). | 1.5.1.26 Beta-alanopine dehydrogenase. |

| | |
|---|---|
| 1.5.1.27 1,2-dehydroreticulinium reductase (NADPH). | 1.6.5.7 2-hydroxy-1,4-benzoquinone reductase. |
| 1.5.1.28 Opine dehydrogenase. | 1.6.6.9 Trimethylamine-N-oxide reductase. |
| 1.5.1.29 FMN reductase. | 1.6.99.1 NADPH dehydrogenase. |
| 1.5.1.30 Flavin reductase. | 1.6.99.2 NAD(P)H dehydrogenase (quinone). |
| 1.5.1.31 Berberine reductase. | 1.6.99.3 NADH dehydrogenase. |
| 1.5.1.32 Vomilenine reductase. | 1.6.99.5 NADH dehydrogenase (quinone). |
| 1.5.1.33 Pteridine reductase. | 1.6.99.6 NADPH dehydrogenase (quinone). |
| 1.5.1.34 6,7-dihydropteridine reductase. | 1.7.1.1 Nitrate reductase (NADH). |
| 1.5.3.1 Sarcosine oxidase. | 1.7.1.2 Nitrate reductase (NAD(P)H). |
| 1.5.3.2 N-methyl-L-amino-acid oxidase. | 1.7.1.3 Nitrate reductase (NADPH). |
| 1.5.3.4 N(6)-methyl-lysine oxidase. | 1.7.1.4 Nitrite reductase (NAD(P)H). |
| 1.5.3.5 (S)-6-hydroxynicotine oxidase. | 1.7.1.5 Hyponitrite reductase. |
| 1.5.3.6 (R)-6-hydroxynicotine oxidase. | 1.7.1.6 Azobenzene reductase. |
| 1.5.3.7 L-pipecolate oxidase. | 1.7.1.7 GMP reductase. |
| 1.5.3.10 Dimethylglycine oxidase. | 1.7.1.9 Nitroquinoline-N-oxide reductase. |
| 1.5.3.11 Polyamine oxidase. | 1.7.1.10 Hydroxylamine reductase (NADH). |
| 1.5.3.12 Dihydrobenzophenanthridine oxidase. | 1.7.1.11 4-(dimethylamino)phenylazoxybenzene reductase. |
| 1.5.4.1 Pyrimidodiazepine synthase. | 1.7.1.12 N-hydroxy-2-acetamidofluorene reductase. |
| 1.5.5.1 Electron-transferring-flavoprotein dehydrogenase. | 1.7.2.1 Nitrite reductase (NO-forming). |
| 1.5.8.1 Dimethylamine dehydrogenase. | 1.7.2.2 Nitrite reductase (cytochrome; ammonia-forming). |

| | |
|---|---|
| 1.5.8.2 Trimethylamine dehydrogenase. | 1.7.2.3 Trimethylamine-N-oxide reductase (cytochrome c). |
| 1.5.99.1 Sarcosine dehydrogenase. | 1.7.3.1 Nitroethane oxidase. |
| 1.5.99.2 Dimethylglycine dehydrogenase. | 1.7.3.2 Acetylindoxyl oxidase. |
| 1.5.99.3 L-pipecolate dehydrogenase. | 1.7.3.3 Urate oxidase. |
| 1.5.99.4 Nicotine dehydrogenase. | 1.7.3.4 Hydroxylamine oxidase. |
| 1.5.99.5 Methylglutamate dehydrogenase. | 1.7.3.5 3-aci-nitropropanoate oxidase. |
| 1.5.99.6 Spermidine dehydrogenase. | 1.7.7.1 Ferredoxin--nitrite reductase. |
| 1.5.99.8 Proline dehydrogenase. | 1.7.7.2 Ferredoxin--nitrate reductase. |
| 1.5.99.9 Methylenetetrahydromethanopterin dehydrogenase. | 1.7.99.1 Hydroxylamine reductase. |
| 1.5.99.11 5,10-methylenetetrahydromethanopterin reductase. | 1.7.99.4 Nitrate reductase. |
| 1.5.99.12 Cytokinin dehydrogenase. | 1.7.99.5 5,10-methylenetetrahydrofolate reductase (FADH(2)). |
| 1.6.1.1 NAD(P)(+) transhydrogenase (B-specific). | 1.7.99.6 Nitrous-oxide reductase. |
| 1.6.1.2 NAD(P)(+) transhydrogenase (AB-specific). | 1.7.99.7 Nitric-oxide reductase. |
| 1.6.2.2 Cytochrome-b5 reductase. | 1.7.99.8 Hydroxylamine oxidoreductase. |
| 1.6.2.4 NADPH--hemoprotein reductase. | 1.8.1.2 Sulfite reductase (NADPH). |
| 1.6.2.5 NADPH--cytochrome-c2 reductase. | 1.8.1.3 Hypotaurine dehydrogenase. |
| 1.6.2.6 Leghemoglobin reductase. | 1.8.1.4 Dihydrolipoyl dehydrogenase. |
| 1.6.3.1 NAD(P)H oxidase. | 1.8.1.5 2-oxopropyl-CoM reductase (carboxylating). |
| 1.6.5.3 NADH dehydrogenase (ubiquinone). | 1.8.1.6 Cystine reductase. |
| 1.6.5.4 Monodehydroascorbate reductase (NADH). | 1.8.1.7 Glutathione-disulfide reductase. |
| 1.6.5.5 NADPH:quinone reductase. | 1.8.1.8 Protein-disulfide reductase. |
| 1.6.5.6 p-benzoquinone reductase (NADPH). | 1.8.1.9 Thioredoxin-disulfide reductase. |

| | |
|---|---|
| 1.8.1.10 CoA-glutathione reductase. | 1.11.1.3 Fatty-acid peroxidase. |
| 1.8.1.11 Asparagusate reductase. | 1.11.1.5 Cytochrome-c peroxidase. |
| 1.8.1.12 Trypanothione-disulfide reductase. | 1.11.1.6 Catalase. |
| 1.8.1.13 Bis-gamma-glutamylcystine reductase. | 1.11.1.7 Peroxidase. |
| 1.8.1.14 CoA-disulfide reductase. | 1.11.1.8 Iodide peroxidase. |
| 1.8.1.15 Mycothione reductase. | 1.11.1.9 Glutathione peroxidase. |
| 1.8.2.1 Sulfite dehydrogenase. | 1.11.1.10 Chloride peroxidase. |
| 1.8.2.2 Thiosulfate dehydrogenase. | 1.11.1.11 L-ascorbate peroxidase. |
| 1.8.3.1 Sulfite oxidase. | 1.11.1.12 Phospholipid-hydroperoxide glutathione peroxidase. |
| 1.8.3.2 Thiol oxidase. | 1.11.1.13 Manganese peroxidase. |
| 1.8.3.3 Glutathione oxidase. | 1.11.1.14 Diarylpropane peroxidase. |
| 1.8.3.4 Methanethiol oxidase. | 1.12.1.2 Hydrogen dehydrogenase. |
| 1.8.3.5 Prenylcysteine oxidase. | 1.12.1.3 Hydrogen dehydrogenase (NADP+). |
| 1.8.4.1 Glutathione--homocystine transhydrogenase. | 1.12.2.1 Cytochrome-c3 hydrogenase. |
| 1.8.4.2 Protein-disulfide reductase (glutathione). | 1.12.5.1 Hydrogen:quinone oxidoreductase. |
| 1.8.4.3 Glutathione--CoA-glutathione transhydrogenase. | 1.12.7.2 Ferredoxin hydrogenase. |
| 1.8.4.4 Glutathione--cystine transhydrogenase. | 1.12.98.1 Coenzyme F420 hydrogenase. |
| 1.8.4.5 Methionine-S-oxide reductase. | 1.12.98.2 5,10-methenyltetrahydromethanopterin hydrogenase. |
| 1.8.4.6 Protein-methionine-S-oxide reductase. | 1.12.98.3 Methanosarcina-phenazine hydrogenase. |

| | |
|---|---|
| 1.8.4.7 Enzyme-thiol transhydrogenase (glutathione-disulfide). | 1.12.99.6 Hydrogenase (acceptor). |
| 1.8.4.8 Phosphoadenylyl-sulfate reductase (thioredoxin). | 1.13.11.1 Catechol 1,2-dioxygenase. |
| 1.8.4.9 Adenylyl-sulfate reductase (glutathione). | 1.13.11.2 Catechol 2,3-dioxygenase. |
| 1.8.4.10 Adenylyl-sulfate reductase (thioredoxin). | 1.13.11.3 Protocatechuate 3,4-dioxygenase. |
| 1.8.5.1 Glutathione dehydrogenase (ascorbate). | 1.13.11.4 Gentisate 1,2-dioxygenase. |
| 1.8.7.1 Sulfite reductase (ferredoxin). | 1.13.11.5 Homogentisate 1,2-dioxygenase. |
| 1.8.98.1 CoB--CoM heterodisulfide reductase. | 1.13.11.6 3-hydroxyanthranilate 3,4-dioxygenase. |
| 1.8.99.1 Sulfite reductase. | 1.13.11.8 Protocatechuate 4,5-dioxygenase. |
| 1.8.99.2 Adenylyl-sulfate reductase. | 1.13.11.9 2,5-dihydroxypyridine 5,6-dioxygenase. |
| 1.8.99.3 Hydrogensulfite reductase. | 1.13.11.10 7,8-dihydroxykynurenate 8,8a-dioxygenase. |
| 1.9.3.1 Cytochrome-c oxidase. | 1.13.11.11 Tryptophan 2,3-dioxygenase. |
| 1.9.6.1 Nitrate reductase (cytochrome). | 1.13.11.12 Lipoxygenase. |
| 1.9.99.1 Iron--cytochrome-c reductase. | 1.13.11.13 Ascorbate 2,3-dioxygenase. |
| 1.10.1.1 Trans-acenaphthene-1,2-diol dehydrogenase. | 1.13.11.14 2,3-dihydroxybenzoate 3,4-dioxygenase. |
| 1.10.2.1 L-ascorbate--cytochrome-b5 reductase. | 1.13.11.15 3,4-dihydroxyphenylacetate 2,3-dioxygenase. |
| 1.10.2.2 Ubiquinol--cytochrome-c reductase. | 1.13.11.16 3-carboxyethylcatechol 2,3-dioxygenase. |
| 1.10.3.1 Catechol oxidase. | 1.13.11.17 Indole 2,3-dioxygenase. |
| 1.10.3.2 Laccase. | 1.13.11.18 Sulfur dioxygenase. |
| 1.10.3.3 L-ascorbate oxidase. | 1.13.11.19 Cysteamine dioxygenase. |
| 1.10.3.4 O-aminophenol oxidase. | 1.13.11.20 Cysteine dioxygenase. |
| 1.10.3.5 3-hydroxyanthranilate oxidase. | 1.13.11.22 Caffeate 3,4-dioxygenase. |
| 1.10.3.6 Rifamycin-B oxidase. | 1.13.11.23 2,3-dihydroxyindole 2,3-dioxygenase. |
| 1.10.99.1 Plastoquinol--plastocyanin reductase. | 1.13.11.24 Quercetin 2,3-dioxygenase. |
| 1.11.1.1 NADH peroxidase. | 1.13.11.25 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione 4,5-dioxygenase. |
| 1.11.1.2 NADPH peroxidase. | 1.13.11.26 Peptide-tryptophan 2,3-dioxygenase. |

| | |
|---|---|
| 1.13.11.27 4-hydroxyphenylpyruvate dioxygenase. | 1.14.11.4 Procollagen-lysine 5-dioxygenase. |
| 1.13.11.28 2,3-dihydroxybenzoate 2,3-dioxygenase. | 1.14.11.6 Thymine dioxygenase. |
| 1.13.11.29 Stizolobate synthase. | 1.14.11.7 Procollagen-proline 3-dioxygenase. |
| 1.13.11.30 Stizolobinate synthase. | 1.14.11.8 Trimethyllysine dioxygenase. |
| 1.13.11.31 Arachidonate 12-lipoxygenase. | 1.14.11.9 Naringenin 3-dioxygenase. |
| 1.13.11.32 2-nitropropane dioxygenase. | 1.14.11.10 Pyrimidine-deoxynucleoside 1'-dioxygenase. |
| 1.13.11.33 Arachidonate 15-lipoxygenase. | 1.14.11.11 Hyoscyamine (6S)-dioxygenase. |
| 1.13.11.34 Arachidonate 5-lipoxygenase. | 1.14.11.12 Gibberellin-44 dioxygenase. |
| 1.13.11.35 Pyrogallol 1,2-oxygenase. | 1.14.11.13 Gibberellin 2-beta-dioxygenase. |
| 1.13.11.36 Chloridazon-catechol dioxygenase. | 1.14.11.14 6-beta-hydroxyhyoscyamine epoxidase. |
| 1.13.11.37 Hydroxyquinol 1,2-dioxygenase. | 1.14.11.15 Gibberellin 3-beta-dioxygenase. |
| 1.13.11.38 1-hydroxy-2-naphthoate 1,2-dioxygenase. | 1.14.11.16 Peptide-aspartate beta-dioxygenase. |
| 1.13.11.39 Biphenyl-2,3-diol 1,2-dioxygenase. | 1.14.11.17 Taurine dioxygenase. |
| 1.13.11.40 Arachidonate 8-lipoxygenase. | 1.14.11.18 Phytanoyl-CoA dioxygenase. |
| 1.13.11.41 2,4'-dihydroxyacetophenone dioxygenase. | 1.14.11.19 Leucocyanidin oxygenase. |
| 1.13.11.42 Indoleamine-pyrrole 2,3-dioxygenase. | 1.14.11.20 Desacetoxyvindoline 4-hydroxylase. |

| | |
|---|---|
| 1.13.11.43 Lignostilbene alpha-beta-dioxygenase. | 1.14.11.21 Clavaminate synthase. |
| 1.13.11.44 Linoleate diol synthase. | 1.14.12.1 Anthranilate 1,2-dioxygenase (deaminating, decarboxylating). |
| 1.13.11.45 Linoleate 11-lipoxygenase. | 1.14.12.3 Benzene 1,2-dioxygenase. |
| 1.13.11.46 4-hydroxymandelate synthase. | 1.14.12.4 3-hydroxy-2-methylpyridinecarboxylate dioxygenase. |
| 1.13.11.47 3-hydroxy-4-oxoquinoline 2,4-dioxygenase. | 1.14.12.5 5-pyridoxate dioxygenase. |
| 1.13.11.48 3-hydroxy-2-methylquinolin-4-one 2,4-dioxygenase. | 1.14.12.7 Phthalate 4,5-dioxygenase. |
| 1.13.11.49 Chlorite O(2)-lyase. | 1.14.12.8 4-sulfobenzoate 3,4-dioxygenase. |
| 1.13.11.50 Acetylacetone-cleaving enzyme. | 1.14.12.9 4-chlorophenylacetate 3,4-dioxygenase. |
| 1.13.12.1 Arginine 2-monooxygenase. | 1.14.12.10 Benzoate 1,2-dioxygenase. |
| 1.13.12.2 Lysine 2-monooxygenase. | 1.14.12.11 Toluene dioxygenase. |
| 1.13.12.3 Tryptophan 2-monooxygenase. | 1.14.12.12 Naphthalene 1,2-dioxygenase. |
| 1.13.12.4 Lactate 2-monooxygenase. | 1.14.12.13 2-chlorobenzoate 1,2-dioxygenase. |
| 1.13.12.5 Renilla-luciferin 2-monooxygenase. | 1.14.12.14 2-aminobenzenesulfonate 2,3-dioxygenase. |
| 1.13.12.6 Cypridina-luciferin 2-monooxygenase. | 1.14.12.15 Terephthalate 1,2-dioxygenase. |
| 1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolyzing). | 1.14.12.16 2-hydroxyquinoline 5,6-dioxygenase. |
| 1.13.12.8 Watasenia-luciferin 2-monooxygenase. | 1.14.12.17 Nitric oxide dioxygenase. |
| 1.13.12.9 Phenylalanine 2-monooxygenase. | 1.14.12.18 Biphenyl 2,3-dioxygenase. |
| 1.13.12.11 Methylphenyltetrahydropyridine N-monooxygenase. | 1.14.13.1 Salicylate 1-monooxygenase. |
| 1.13.12.12 Apo-beta-carotenoid-14',13'-dioxygenase. | 1.14.13.2 4-hydroxybenzoate 3-monooxygenase. |
| 1.13.12.13 Oplophorus-luciferin 2-monooxygenase. | 1.14.13.3 4-hydroxyphenylacetate 3-monooxygenase. |
| 1.13.99.1 Inositol oxygenase. | 1.14.13.4 Melilotate 3-monooxygenase. |
| 1.13.99.3 Tryptophan 2'-dioxygenase. | 1.14.13.5 Imidazoleacetate 4-monooxygenase. |
| 1.14.11.1 Gamma-butyrobetaine dioxygenase. | 1.14.13.6 Orcinol 2-monooxygenase. |
| 1.14.11.2 Procollagen-proline dioxygenase. | 1.14.13.7 Phenol 2-monooxygenase. |
| 1.14.11.3 Pyrimidine-deoxynucleoside 2'-dioxygenase. | 1.14.13.8 Dimethylaniline monooxygenase (N-oxide-forming). |

| | |
|---|---|
| 1.14.16.4 Tryptophan 5-monooxygenase. | 1.14.99.33 Delta(12)-fatty acid dehydrogenase. |
| 1.14.16.5 Glyceryl-ether monooxygenase. | 1.14.99.34 Monoprenyl isoflavone epoxidase. |
| 1.14.16.6 Mandelate 4-monooxygenase. | 1.14.99.35 Thiophene-2-carbonyl-CoA monooxygenase. |
| 1.14.17.1 Dopamine beta-monooxygenase. | 1.14.99.36 Beta-carotene 15,15'-monooxygenase. |
| 1.14.17.3 Peptidylglycine monooxygenase. | 1.14.99.37 Taxadiene 5-alpha-hydroxylase. |
| 1.14.17.4 Aminocyclopropanecarboxylate oxidase. | 1.15.1.1 Superoxide dismutase. |
| 1.14.18.1 Monophenol monooxygenase. | 1.15.1.2 Superoxide reductase. |
| 1.14.18.2 CMP-N-acetylneuraminate monooxygenase. | 1.16.1.1 Mercury(II) reductase. |
| 1.14.19.1 Stearoyl-CoA 9-desaturase. | 1.16.1.2 Diferric-transferrin reductase. |
| 1.14.19.2 Acyl-[acyl-carrier-protein] desaturase. | 1.16.1.3 Aquacobalamin reductase. |
| 1.14.19.3 Linoleoyl-CoA desaturase. | 1.16.1.4 Cob(II)alamin reductase. |
| 1.14.20.1 Deacetoxycephalosporin-C synthase. | 1.16.1.5 Aquacobalamin reductase (NADPH). |

| | |
|---|---|
| 1.14.21.1 (S)-stylopine synthase. | 1.16.1.6 Cyanocobalamin reductase (cyanide-eliminating). |
| 1.14.21.2 (S)-cheilanthifoline synthase. | 1.16.1.7 Ferric-chelate reductase. |
| 1.14.21.3 Berbamunine synthase. | 1.16.1.8 [Methionine synthase] reductase. |
| 1.14.21.4 Salutaridine synthase. | 1.16.3.1 Ferroxidase. |
| 1.14.21.5 (S)-canadine synthase. | 1.16.8.1 Cob(II)yrinic acid a,c-diamide reductase. |
| 1.14.99.1 Prostaglandin-endoperoxide synthase. | 1.17.1.1 CDP-4-dehydro-6-deoxyglucose reductase. |
| 1.14.99.2 Kynurenine 7,8-hydroxylase. | 1.17.1.2 4-hydroxy-3-methylbut-2-enyl diphosphate reductase. |
| 1.14.99.3 Heme oxygenase (decyclizing). | 1.17.1.3 Leucoanthocyanidin reductase. |
| 1.14.99.4 Progesterone monooxygenase. | 1.17.1.4 Xanthine dehydrogenase. |
| 1.14.99.7 Squalene monooxygenase. | 1.17.1.5 Nicotinate dehydrogenase. |
| 1.14.99.9 Steroid 17-alpha-monooxygenase. | 1.17.3.1 Pteridine oxidase. |
| 1.14.99.10 Steroid 21-monooxygenase. | 1.17.3.2 Xanthine oxidase. |
| 1.14.99.11 Estradiol 6-beta-monooxygenase. | 1.17.3.3 6-hydroxynicotinate dehydrogenase. |
| 1.14.99.12 Androst-4-ene-3,17-dione monooxygenase. | 1.17.4.1 Ribonucleoside-diphosphate reductase. |
| 1.14.99.14 Progesterone 11-alpha-monooxygenase. | 1.17.4.2 Ribonucleoside-triphosphate reductase. |
| 1.14.99.15 4-methoxybenzoate monooxygenase (O-demethylating). | 1.17.4.3 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase. |
| 1.14.99.19 Plasmanylethanolamine desaturase. | 1.17.5.1 Phenylacetyl-CoA dehydrogenase. |
| 1.14.99.20 Phylloquinone monooxygenase (2,3-epoxidizing). | 1.17.99.1 4-cresol dehydrogenase (hydroxylating). |
| 1.14.99.21 Latia-luciferin monooxygenase (demethylating). | 1.17.99.2 Ethylbenzene hydroxylase. |
| 1.14.99.22 Ecdysone 20-monooxygenase. | 1.18.1.1 Rubredoxin--NAD(+) reductase. |
| 1.14.99.23 3-hydroxybenzoate 2-monooxygenase. | 1.18.1.2 Ferredoxin--NADP(+) reductase. |
| 1.14.99.24 Steroid 9-alpha-monooxygenase. | 1.18.1.3 Ferredoxin--NAD(+) reductase. |
| 1.14.99.26 2-hydroxypyridine 5-monooxygenase. | 1.18.1.4 Rubredoxin--NAD(P)(+) reductase. |
| 1.14.99.27 Juglone 3-monooxygenase. | 1.18.6.1 Nitrogenase. |
| 1.14.99.28 Linalool 8-monooxygenase. | 1.19.6.1 Nitrogenase (flavodoxin). |
| 1.14.99.29 Deoxyhypusine monooxygenase. | 1.20.1.1 Phosphonate dehydrogenase. |
| 1.14.99.30 Carotene 7,8-desaturase. | 1.20.4.1 Arsenate reductase (glutaredoxin). |
| 1.14.99.31 Myristoyl-CoA 11-(E) desaturase. | 1.20.4.2 Methylarsonate reductase. |
| 1.14.99.32 Myristoyl-CoA 11-(Z) desaturase. | 1.20.98.1 Arsenate reductase (azurin). |

| | |
|---|---|
| 1.20.99.1 Arsenate reductase (donor). | 2.1.1.15 Fatty-acid O-methyltransferase. |
| 1.21.3.1 Isopenicillin-N synthase. | 2.1.1.16 Methylene-fatty-acyl-phospholipid synthase. |
| 1.21.3.2 Columbamine oxidase. | 2.1.1.17 Phosphatidylethanolamine N-methyltransferase. |
| 1.21.3.3 Reticuline oxidase. | 2.1.1.18 Polysaccharide O-methyltransferase. |
| 1.21.3.4 Sulochrin oxidase ((+)-bisdechlorogeodin-forming). | 2.1.1.19 Trimethylsulfonium--tetrahydrofolate N-methyltransferase. |
| 1.21.3.5 Sulochrin oxidase ((-)-bisdechlorogeodin-forming). | 2.1.1.20 Glycine N-methyltransferase. |
| 1.21.3.6 Aureusidin synthase. | 2.1.1.21 Methylamine--glutamate N-methyltransferase. |
| 1.21.4.1 D-proline reductase (dithiol). | 2.1.1.22 Carnosine N-methyltransferase. |

| | |
|---|---|
| 1.21.4.2 Glycine reductase. | 2.1.1.25 Phenol O-methyltransferase. |
| 1.21.4.3 Sarcosine reductase. | 2.1.1.26 Iodophenol O-methyltransferase. |
| 1.21.4.4 Betaine reductase. | 2.1.1.27 Tyramine N-methyltransferase. |
| 1.21.99.1 Beta-cyclopiazonate dehydrogenase. | 2.1.1.28 Phenylethanolamine N-methyltransferase. |
| 1.97.1.1 Chlorate reductase. | 2.1.1.29 tRNA (cytosine-5-)-methyltransferase. |
| 1.97.1.2 Pyrogallol hydroxytransferase. | 2.1.1.31 tRNA (guanine-N(1)-)-methyltransferase. |
| 1.97.1.3 Sulfur reductase. | 2.1.1.32 tRNA (guanine-N(2)-)-methyltransferase. |
| 1.97.1.4 [Formate acetyltransferase] activating enzyme. | 2.1.1.33 tRNA (guanine-N(7)-)-methyltransferase. |
| 1.97.1.8 Tetrachloroethene reductive dehalogenase. | 2.1.1.34 tRNA (guanosine-2'-O-)-methyltransferase. |
| 1.97.1.9 Selenate reductase. | 2.1.1.35 tRNA (uracil-5-)-methyltransferase. |
| 1.97.1.10 Thyroxine 5'-deiodinase. | 2.1.1.36 tRNA (adenine-N(1)-)-methyltransferase. |
| 1.97.1.11 Thyroxine 5-deiodinase. | 2.1.1.37 DNA (cytosine-5-)-methyltransferase. |
| | 2.1.1.38 O-demethylpuromycin O-methyltransferase. |
| ENZYME: 2.-.-.- | 2.1.1.39 Inositol 3-methyltransferase. |
| 2.1.1.1 Nicotinamide N-methyltransferase. | 2.1.1.40 Inositol 1-methyltransferase. |
| 2.1.1.2 Guanidinoacetate N-methyltransferase. | 2.1.1.41 Sterol 24-C-methyltransferase. |
| 2.1.1.3 Thetin--homocysteine S-methyltransferase. | 2.1.1.42 Luteolin O-methyltransferase. |
| 2.1.1.4 Acetylserotonin O-methyltransferase. | 2.1.1.43 Histone-lysine N-methyltransferase. |
| 2.1.1.5 Betaine--homocysteine S-methyltransferase. | 2.1.1.44 Dimethylhistidine N-methyltransferase. |
| 2.1.1.6 Catechol O-methyltransferase. | 2.1.1.45 Thymidylate synthase. |
| 2.1.1.7 Nicotinate N-methyltransferase. | 2.1.1.46 Isoflavone 4'-O-methyltransferase. |
| 2.1.1.8 Histamine N-methyltransferase. | 2.1.1.47 Indolepyruvate C-methyltransferase. |
| 2.1.1.9 Thiol S-methyltransferase. | 2.1.1.48 rRNA (adenine-N(6)-)-methyltransferase. |
| 2.1.1.10 Homocysteine S-methyltransferase. | 2.1.1.49 Amine N-methyltransferase. |
| 2.1.1.11 Magnesium protoporphyrin IX methyltransferase. | 2.1.1.50 Loganate O-methyltransferase. |
| 2.1.1.12 Methionine S-methyltransferase. | 2.1.1.51 rRNA (guanine-N(1)-)-methyltransferase. |
| 2.1.1.13 Methionine synthase. | 2.1.1.52 rRNA (guanine-N(2)-)-methyltransferase. |
| 2.1.1.14 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase. | 2.1.1.53 Putrescine N-methyltransferase. |

| | |
|---|---|
| 2.1.1.54 Deoxycytidylate C-methyltransferase. | 2.1.1.86 Tetrahydromethanopterin S-methyltransferase. |
| 2.1.1.55 tRNA (adenine-N(6)-)-methyltransferase. | 2.1.1.87 Pyridine N-methyltransferase. |
| 2.1.1.56 mRNA (guanine-N(7)-)-methyltransferase. | 2.1.1.88 8-hydroxyquercitin 8-O-methyltransferase. |
| 2.1.1.57 mRNA (nucleoside-2'-O-)-methyltransferase. | 2.1.1.89 Tetrahydrocolumbamine 2-O-methyltransferase. |
| 2.1.1.59 [Cytochrome c]-lysine N-methyltransferase. | 2.1.1.90 Methanol--5-hydroxybenzimidazolylcobamide Co-methyltransferase. |

| | |
|---|---|
| 2.1.1.60 Calmodulin-lysine N-methyltransferase. | 2.1.1.91 Isobutyraldoxime O-methyltransferase. |
| 2.1.1.61 tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase. | 2.1.1.92 Bergaptol O-methyltransferase. |
| 2.1.1.62 mRNA (2'-O-methyladenosine-N(6)-)-methyltransferase. | 2.1.1.93 Xanthotoxol O-methyltransferase. |
| 2.1.1.63 Methylated-DNA--[protein]-cysteine S-methyltransferase. | 2.1.1.94 11-O-demethyl-17-O-deacetylvindoline O-methyltransferase. |
| 2.1.1.64 3-demethylubiquinone-9 3-O-methyltransferase. | 2.1.1.95 Tocopherol O-methyltransferase. |
| 2.1.1.65 Licodione 2'-O-methyltransferase. | 2.1.1.96 Thioether S-methyltransferase. |
| 2.1.1.66 rRNA (adenosine-2'-O-)-methyltransferase. | 2.1.1.97 3-hydroxyanthranilate 4-C-methyltransferase. |
| 2.1.1.67 Thiopurine S-methyltransferase. | 2.1.1.98 Diphthine synthase. |
| 2.1.1.68 Caffeate O-methyltransferase. | 2.1.1.99 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase. |
| 2.1.1.69 5-hydroxyfuranocoumarin 5-O-methyltransferase. | 2.1.1.100 Protein-S-isoprenylcysteine O-methyltransferase. |
| 2.1.1.70 8-hydroxyfuranocoumarin 8-O-methyltransferase. | 2.1.1.101 Macrocin O-methyltransferase. |
| 2.1.1.71 Phosphatidyl-N-methylethanolamine N-methyltransferase. | 2.1.1.102 Demethylmacrocin O-methyltransferase. |
| 2.1.1.72 Site-specific DNA-methyltransferase (adenine-specific). | 2.1.1.103 Phosphoethanolamine N-methyltransferase. |
| 2.1.1.74 Methylenetetrahydrofolate--tRNA-(uracil-5-)-methyltransferase (FADH(2)-oxidizing). | 2.1.1.104 Caffeoyl-CoA O-methyltransferase. |
| 2.1.1.75 Apigenin 4'-O-methyltransferase. | 2.1.1.105 N-benzoyl-4-hydroxyanthranilate 4-O-methyltransferase. |
| 2.1.1.76 Quercetin 3-O-methyltransferase. | 2.1.1.106 Tryptophan 2-C-methyltransferase. |
| 2.1.1.77 Protein-L-isoaspartate(D-aspartate) O-methyltransferase. | 2.1.1.107 Uroporphyrin-III C-methyltransferase. |
| 2.1.1.78 Isoorientin 3'-O-methyltransferase. | 2.1.1.108 6-hydroxymellein O-methyltransferase. |
| 2.1.1.79 Cyclopropane-fatty-acyl-phospholipid synthase. | 2.1.1.109 Demethylsterigmatocystin 6-O-methyltransferase. |
| 2.1.1.80 Protein-glutamate O-methyltransferase. | 2.1.1.110 Sterigmatocystin 7-O-methyltransferase. |
| 2.1.1.82 3-methylquercitin 7-O-methyltransferase. | 2.1.1.111 Anthranilate N-methyltransferase. |
| 2.1.1.83 3,7-dimethylquercitin 4'-O-methyltransferase. | 2.1.1.112 Glucuronoxylan 4-O-methyltransferase. |
| 2.1.1.84 Methylquercetagetin 6-O-methyltransferase. | 2.1.1.113 Site-specific DNA-methyltransferase (cytosine-N(4)-specific). |
| 2.1.1.85 Protein-histidine N-methyltransferase. | 2.1.1.114 Hexaprenyldihydroxybenzoate methyltransferase. |

| | |
|---|---|
| 2.1.1.115 (RS)-1-benzyl-1,2,3,4-tetrahydroisoquinoline N-methyltransferase. | 2.1.2.1 Glycine hydroxymethyltransferase. |
| 2.1.1.116 3'-hydroxy-N-methyl-(S)-coclaurine 4'-O-methyltransferase. | 2.1.2.2 Phosphoribosylglycinamide formyltransferase. |
| 2.1.1.117 (S)-scoulerine 9-O-methyltransferase. | 2.1.2.3 Phosphoribosylaminoimidazolecarboxamide formyltransferase. |
| 2.1.1.118 Columbamine O-methyltransferase. | 2.1.2.4 Glycine formimidoyltransferase. |
| 2.1.1.119 10-hydroxydihydrosanguinarine 10-O-methyltransferase. | 2.1.2.5 Glutamate formimidoyltransferase. |

| | |
|---|---|
| 2.1.1.120 12-hydroxydihydrochelirubine 12-O-methyltransferase. | 2.1.2.7 D-alanine 2-hydroxymethyltransferase. |
| 2.1.1.121 6-O-methylnorlaudanosoline 5'-O-methyltransferase. | 2.1.2.8 Deoxycytidylate 5-hydroxymethyltransferase. |
| 2.1.1.122 (S)-tetrahydroprotoberberine N-methyltransferase. | 2.1.2.9 Methionyl-tRNA formyltransferase. |
| 2.1.1.123 [Cytochrome-c]-methionine S-methyltransferase. | 2.1.2.10 Aminomethyltransferase. |
| 2.1.1.124 [Cytochrome-c]-arginine N-methyltransferase. | 2.1.2.11 3-methyl-2-oxobutanoate hydroxymethyltransferase. |
| 2.1.1.125 Histone-arginine N-methyltransferase. | 2.1.3.1 Methylmalonyl-CoA carboxytransferase. |
| 2.1.1.126 [Myelin basic protein]-arginine N-methyltransferase. | 2.1.3.2 Aspartate carbamoyltransferase. |
| 2.1.1.127 [Ribulose-bisphosphate carboxylase]-lysine N-methyltransferase. | 2.1.3.3 Ornithine carbamoyltransferase. |
| 2.1.1.128 (RS)-norcoclaurine 6-O-methyltransferase. | 2.1.3.5 Oxamate carbamoyltransferase. |
| 2.1.1.129 Inositol 4-methyltransferase. | 2.1.3.6 Putrescine carbamoyltransferase. |
| 2.1.1.130 Precorrin-2 C(20)-methyltransferase. | 2.1.3.7 3-hydroxymethylcephem carbamoyltransferase. |
| 2.1.1.131 Precorrin-3B C(17)-methyltransferase. | 2.1.3.8 Lysine carbamoyltransferase. |
| 2.1.1.132 Precorrin-6Y C(5,15)-methyltransferase (decarboxylating). | 2.1.4.1 Glycine amidinotransferase. |
| 2.1.1.133 Precorrin-4 C(11)-methyltransferase. | 2.1.4.2 Scyllo-inosamine-4-phosphate amidinotransferase. |
| 2.1.1.136 Chlorophenol O-methyltransferase. | 2.2.1.1 Transketolase. |
| 2.1.1.137 Arsenite methyltransferase. | 2.2.1.2 Transaldolase. |
| 2.1.1.139 3'-demethylstaurosporine O-methyltransferase. | 2.2.1.3 Formaldehyde transketolase. |
| 2.1.1.140 (S)-coclaurine-N-methyltransferase. | 2.2.1.4 Acetoin--ribose-5-phosphate transaldolase. |
| 2.1.1.141 Jasmonate O-methyltransferase. | 2.2.1.5 2-hydroxy-3-oxoadipate synthase. |
| 2.1.1.142 Cycloartenol 24-C-methyltransferase. | 2.2.1.6 Acetolactate synthase. |
| 2.1.1.143 24-methylenesterol C-methyltransferase. | 2.2.1.7 1-deoxy-D-xylulose-5-phosphate synthase. |
| 2.1.1.144 Trans-aconitate 2-methyltransferase. | 2.2.1.8 Fluorothreonine transaldolase. |
| 2.1.1.145 Trans-aconitate 3-methyltransferase. | 2.3.1.1 Amino-acid N-acetyltransferase. |
| 2.1.1.146 (Iso)eugenol O-methyltransferase. | 2.3.1.2 Imidazole N-acetyltransferase. |
| 2.1.1.147 Corydaline synthase. | 2.3.1.3 Glucosamine N-acetyltransferase. |
| 2.1.1.148 Thymidylate synthase (FAD). | 2.3.1.4 Glucosamine 6-phosphate N-acetyltransferase. |
| 2.1.1.149 Myricetin O-methyltransferase. | 2.3.1.5 Arylamine N-acetyltransferase. |
| 2.1.1.150 Isoflavone 7-O-methyltransferase. | 2.3.1.6 Choline O-acetyltransferase. |
| 2.1.1.151 Cobalt-factor II C(20)-methyltransferase. | 2.3.1.7 Carnitine O-acetyltransferase. |
| 2.1.1.152 Precorrin-6A synthase (deacetylating). | 2.3.1.8 Phosphate acetyltransferase. |
| 2.3.1.9 Acetyl-CoA C-acetyltransferase. | 2.3.1.48 Histone acetyltransferase. |
| 2.3.1.10 Hydrogen-sulfide S-acetyltransferase. | 2.3.1.49 Deacetyl-[citrate-(pro-3S)-lyase] S-acetyltransferase. |
| 2.3.1.11 Thioethanolamine S-acetyltransferase. | 2.3.1.50 Serine C-palmitoyltransferase. |
| 2.3.1.12 Dihydrolipoyllysine-residue acetyltransferase. | 2.3.1.51 1-acylglycerol-3-phosphate O-acyltransferase. |
| 2.3.1.13 Glycine N-acyltransferase. | 2.3.1.52 2-acylglycerol-3-phosphate O-acyltransferase. |
| 2.3.1.14 Glutamine N-phenylacetyltransferase. | 2.3.1.53 Phenylalanine N-acetyltransferase. |

| | |
|---|---|
| 2.3.1.15 Glycerol-3-phosphate O-acyltransferase. | 2.3.1.54 Formate C-acetyltransferase. |
| 2.3.1.16 Acetyl-CoA C-acyltransferase. | 2.3.1.56 Aromatic-hydroxylamine O-acetyltransferase. |
| 2.3.1.17 Aspartate N-acetyltransferase. | 2.3.1.57 Diamine N-acetyltransferase. |
| 2.3.1.18 Galactoside O-acetyltransferase. | 2.3.1.58 2,3-diaminopropionate N-oxalyltransferase. |
| 2.3.1.19 Phosphate butyryltransferase. | 2.3.1.59 Gentamicin 2'-N-acetyltransferase. |
| 2.3.1.20 Diacylglycerol O-acyltransferase. | 2.3.1.60 Gentamicin 3'-N-acetyltransferase. |
| 2.3.1.21 Carnitine O-palmitoyltransferase. | 2.3.1.61 Dihydrolipoyllysine-residue succinyltransferase. |
| 2.3.1.22 2-acylglycerol O-acyltransferase. | 2.3.1.62 2-acylglycerophosphocholine O-acyltransferase. |
| 2.3.1.23 1-acylglycerophosphocholine O-acyltransferase. | 2.3.1.63 1-alkylglycerophosphocholine O-acyltransferase. |
| 2.3.1.24 Sphingosine N-acyltransferase. | 2.3.1.64 Agmatine N(4)-coumaroyltransferase. |
| 2.3.1.25 Plasmalogen synthase. | 2.3.1.65 Glycine N-choloyltransferase. |
| 2.3.1.26 Sterol O-acyltransferase. | 2.3.1.66 Leucine N-acetyltransferase. |
| 2.3.1.27 Cortisol O-acetyltransferase. | 2.3.1.67 1-alkylglycerophosphocholine O-acetyltransferase. |
| 2.3.1.28 Chloramphenicol O-acetyltransferase. | 2.3.1.68 Glutamine N-acyltransferase. |
| 2.3.1.29 Glycine C-acetyltransferase. | 2.3.1.69 Monoterpenol O-acetyltransferase. |
| 2.3.1.30 Serine O-acetyltransferase. | 2.3.1.70 CDP-acylglycerol O-arachidonoyltransferase. |
| 2.3.1.31 Homoserine O-acetyltransferase. | 2.3.1.71 Glycine N-benzoyltransferase. |
| 2.3.1.32 Lysine N-acetyltransferase. | 2.3.1.72 Indoleacetylglucose--inositol O-acyltransferase. |
| 2.3.1.33 Histidine N-acetyltransferase. | 2.3.1.73 Diacylglycerol--sterol O-acyltransferase. |
| 2.3.1.34 D-tryptophan N-acetyltransferase. | 2.3.1.74 Naringenin-chalcone synthase. |
| 2.3.1.35 Glutamate N-acetyltransferase. | 2.3.1.75 Long-chain-alcohol O-fatty-acyltransferase. |
| 2.3.1.36 D-amino-acid N-acetyltransferase. | 2.3.1.76 Retinol O-fatty-acyltransferase. |
| 2.3.1.37 5-aminolevulinate synthase. | 2.3.1.77 Triacylglycerol--sterol O-acyltransferase. |
| 2.3.1.38 [Acyl-carrier-protein] S-acetyltransferase. | 2.3.1.78 Heparan-alpha-glucosaminide N-acetyltransferase. |
| 2.3.1.39 [Acyl-carrier-protein] S-malonyltransferase. | 2.3.1.79 Maltose O-acetyltransferase. |
| 2.3.1.40 Acyl-[acyl-carrier-protein]--phospholipid O-acyltransferase. | 2.3.1.80 Cysteine-S-conjugate N-acetyltransferase. |
| 2.3.1.41 3-oxoacyl-[acyl-carrier-protein] synthase. | 2.3.1.81 Aminoglycoside N(3')-acetyltransferase. |
| 2.3.1.42 Glycerone-phosphate O-acyltransferase. | 2.3.1.82 Aminoglycoside N(6')-acetyltransferase. |
| 2.3.1.43 Phosphatidylcholine--sterol O-acyltransferase. | 2.3.1.83 Phosphatidylcholine--dolichol O-acyltransferase. |
| 2.3.1.44 N-acetylneuraminate 4-O-acetyltransferase. | 2.3.1.84 Alcohol O-acetyltransferase. |
| 2.3.1.45 N-acetylneuraminate 7-O(or 9-O)-acetyltransferase. | 2.3.1.85 Fatty-acid synthase. |
| 2.3.1.46 Homoserine O-succinyltransferase. | 2.3.1.86 Fatty-acyl-CoA synthase. |
| 2.3.1.47 8-amino-7-oxononanoate synthase. | 2.3.1.87 Aralkylamine N-acetyltransferase. |
| 2.3.1.88 Peptide alpha-N-acetyltransferase. | 2.3.1.122 Trehalose O-mycolyltransferase. |
| 2.3.1.89 Tetrahydrodipicolinate N-acetyltransferase. | 2.3.1.123 Dolichol O-acyltransferase. |
| 2.3.1.90 Beta-glucogallin O-galloyltransferase. | 2.3.1.125 1-alkyl-2-acetylglycerol O-acyltransferase. |

| | |
|---|---|
| 2.3.1.91 Sinapoylglucose--choline O-sinapoyltransferase. | 2.3.1.126 Isocitrate O-dihydroxycinnamoyltransferase. |
| 2.3.1.92 Sinapoylglucose--malate O-sinapoyltransferase. | 2.3.1.127 Ornithine N-benzoyltransferase. |
| 2.3.1.93 13-hydroxylupinine O-tigloyltransferase. | 2.3.1.128 Ribosomal-protein-alanine N-acetyltransferase. |
| 2.3.1.94 Erythronolide synthase. | 2.3.1.129 Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase. |
| 2.3.1.95 Trihydroxystilbene synthase. | 2.3.1.130 Galactarate O-hydroxycinnamoyltransferase. |
| 2.3.1.96 Glycoprotein N-palmitoyltransferase. | 2.3.1.131 Glucarate O-hydroxycinnamoyltransferase. |
| 2.3.1.97 Glycylpeptide N-tetradecanoyltransferase. | 2.3.1.132 Glucarolactone O-hydroxycinnamoyltransferase. |
| 2.3.1.98 Chlorogenate--glucarate O-hydroxycinnamoyltransferase. | 2.3.1.133 Shikimate O-hydroxycinnamoyltransferase. |
| 2.3.1.99 Quinate O-hydroxycinnamoyltransferase. | 2.3.1.134 Galactolipid O-acyltransferase. |
| 2.3.1.100 Myelin-proteolipid O-palmitoyltransferase. | 2.3.1.135 Phosphatidylcholine--retinol O-acyltransferase. |
| 2.3.1.101 Formylmethanofuran--tetrahydromethanopterin N-formyltransferase. | 2.3.1.136 Polysialic-acid O-acetyltransferase. |
| 2.3.1.102 N(6)-hydroxylysine O-acetyltransferase. | 2.3.1.137 Carnitine O-octanoyltransferase. |
| 2.3.1.103 Sinapoylglucose--sinapoylglucose O-sinapoyltransferase. | 2.3.1.138 Putrescine N-hydroxycinnamoyltransferase. |
| 2.3.1.104 1-alkenylglycerophosphocholine O-acyltransferase. | 2.3.1.139 Ecdysone O-acyltransferase. |
| 2.3.1.105 Alkylglycerophosphate 2-O-acetyltransferase. | 2.3.1.140 Rosmarinate synthase. |
| 2.3.1.106 Tartronate O-hydroxycinnamoyltransferase. | 2.3.1.141 Galactosylacylglycerol O-acyltransferase. |
| 2.3.1.107 17-O-deacetylvindoline O-acetyltransferase. | 2.3.1.142 Glycoprotein O-fatty-acyltransferase. |
| 2.3.1.108 Tubulin N-acetyltransferase. | 2.3.1.143 Beta-glucogallin--tetrakisgalloylglucose O-galloyltransferase. |
| 2.3.1.109 Arginine N-succinyltransferase. | 2.3.1.144 Anthranilate N-benzoyltransferase. |
| 2.3.1.110 Tyramine N-feruloyltransferase. | 2.3.1.145 Piperidine N-piperoyltransferase. |
| 2.3.1.111 Mycocerosate synthase. | 2.3.1.146 Pinosylvin synthase. |
| 2.3.1.112 D-tryptophan N-malonyltransferase. | 2.3.1.147 Glycerophospholipid arachidonoyl-transferase (CoA-independent). |
| 2.3.1.113 Anthranilate N-malonyltransferase. | 2.3.1.148 Glycerophospholipid acyltransferase (CoA-dependent). |
| 2.3.1.114 3,4-dichloroaniline N-malonyltransferase. | 2.3.1.149 Platelet-activating factor acetyltransferase. |
| 2.3.1.115 Isoflavone-7-O-beta-glucoside 6"-O-malonyltransferase. | 2.3.1.150 Salutaridinol 7-O-acetyltransferase. |
| 2.3.1.116 Flavonol-3-O-beta-glucoside O-malonyltransferase. | 2.3.1.151 Benzophenone synthase. |
| 2.3.1.117 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase. | 2.3.1.152 Alcohol O-cinnamoyltransferase. |
| 2.3.1.118 N-hydroxyarylamine O-acetyltransferase. | 2.3.1.153 Anthocyanin 5-aromatic acyltransferase. |
| 2.3.1.119 Icosanoyl-CoA synthase. | 2.3.1.154 Propionyl-CoA C(2)-trimethyltridecanoyltransferase. |
| 2.3.1.121 1-alkenylglycerophosphoethanolamine O-acyltransferase. | 2.3.1.155 Acetyl-CoA C-myristoyltransferase. |

| | |
|---|---|
| 2.3.1.156 Phloroisovalerophenone synthase. | 2.3.3.12 3-propylmalate synthase. |

| | |
|---|---|
| 2.3.1.157 Glucosamine-1-phosphate N-acetyltransferase. | 2.3.3.13 2-isopropylmalate synthase. |
| 2.3.1.158 Phospholipid:diacylglycerol acyltransferase. | 2.3.3.14 Homocitrate synthase. |
| 2.3.1.159 Acridone synthase. | 2.3.3.15 Sulfoacetaldehyde acetyltransferase. |
| 2.3.1.160 Vinorine synthase. | 2.4.1.1 Phosphorylase. |
| 2.3.1.161 Lovastatin nonaketide synthase. | 2.4.1.2 Dextrin dextranase. |
| 2.3.1.162 Taxadien-5-alpha-ol O-acetyltransferase. | 2.4.1.4 Amylosucrase. |
| 2.3.1.163 10-hydroxytaxane O-acetyltransferase. | 2.4.1.5 Dextransucrase. |
| 2.3.1.164 Isopenicillin-N N-acyltransferase. | 2.4.1.7 Sucrose phosphorylase. |
| 2.3.1.165 6-methylsalicylic acid synthase. | 2.4.1.8 Maltose phosphorylase. |
| 2.3.1.166 2-alpha-hydroxytaxane 2-O-benzoyltransferase. | 2.4.1.9 Inulosucrase. |
| 2.3.1.167 10-deacetylbaccatin III 10-O-acetyltransferase. | 2.4.1.10 Levansucrase. |
| 2.3.1.168 Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase. | 2.4.1.11 Glycogen (starch) synthase. |
| 2.3.1.169 CO-methylating acetyl-CoA synthase. | 2.4.1.12 Cellulose synthase (UDP-forming). |
| 2.3.2.1 D-glutamyltransferase. | 2.4.1.13 Sucrose synthase. |
| 2.3.2.2 Gamma-glutamyltransferase. | 2.4.1.14 Sucrose-phosphate synthase. |
| 2.3.2.3 Lysyltransferase. | 2.4.1.15 Alpha, alpha-trehalose-phosphate synthase (UDP-forming). |
| 2.3.2.4 Gamma-glutamylcyclotransferase. | 2.4.1.16 Chitin synthase. |
| 2.3.2.5 Glutaminyl-peptide cyclotransferase. | 2.4.1.17 Glucuronosyltransferase. |
| 2.3.2.6 Leucyltransferase. | 2.4.1.18 1,4-alpha-glucan branching enzyme. |
| 2.3.2.7 Aspartyltransferase. | 2.4.1.19 Cyclomaltodextrin glucanotransferase. |
| 2.3.2.8 Arginyltransferase. | 2.4.1.20 Cellobiose phosphorylase. |
| 2.3.2.9 Agaritine gamma-glutamyltransferase. | 2.4.1.21 Starch synthase. |
| 2.3.2.10 UDP-N-acetylmuramoylpentapeptide-lysine N(6)-alanyltransferase. | 2.4.1.22 Lactose synthase. |
| 2.3.2.11 Alanylphosphatidylglycerol synthase. | 2.4.1.23 Sphingosine beta-galactosyltransferase. |
| 2.3.2.12 Peptidyltransferase. | 2.4.1.24 1,4-alpha-glucan 6-alpha-glucosyltransferase. |
| 2.3.2.13 Protein-glutamine gamma-glutamyltransferase. | 2.4.1.25 4-alpha-glucanotransferase. |
| 2.3.2.14 D-alanine gamma-glutamyltransferase. | 2.4.1.26 DNA alpha-glucosyltransferase. |
| 2.3.2.15 Glutathione gamma-glutamylcysteinyltransferase. | 2.4.1.27 DNA beta-glucosyltransferase. |
| 2.3.3.1 Citrate (Si)-synthase. | 2.4.1.28 Glucosyl-DNA beta-glucosyltransferase. |
| 2.3.3.2 Decylcitrate synthase. | 2.4.1.29 Cellulose synthase (GDP-forming). |
| 2.3.3.3 Citrate (Re)-synthase. | 2.4.1.30 1,3-beta-oligoglucan phosphorylase. |
| 2.3.3.4 Decylhomocitrate synthase. | 2.4.1.31 Laminaribiose phosphorylase. |
| 2.3.3.5 2-methylcitrate synthase. | 2.4.1.32 Glucomannan 4-beta-mannosyltransferase. |
| 2.3.3.6 2-ethylmalate synthase. | 2.4.1.33 Alginate synthase. |
| 2.3.3.7 3-ethylmalate synthase. | 2.4.1.34 1,3-beta-glucan synthase. |
| 2.3.3.8 ATP citrate synthase. | 2.4.1.35 Phenol beta-glucosyltransferase. |
| 2.3.3.9 Malate synthase. | 2.4.1.36 Alpha,alpha-trehalose-phosphate synthase (GDP-forming). |

| | |
|---|---|
| 2.3.3.10 Hydroxymethylglutaryl-CoA synthase. | 2.4.1.37 Fucosylgalactoside 3-alpha-galactosyltransferase. |
| 2.3.3.11 2-hydroxyglutarate synthase. | 2.4.1.38 Beta-N-acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase. |

| | |
|---|---|
| 2.4.1.39 Steroid N-acetylglucosaminyltransferase. | 2.4.1.71 Arylamine glucosyltransferase. |
| 2.4.1.40 Glycoprotein-fucosylgalactoside alpha-N-acetylgalactosaminyltransferase. | 2.4.1.73 Lipopolysaccharide glucosyltransferase II. |
| 2.4.1.41 Polypeptide N-acetylgalactosaminyltransferase. | 2.4.1.74 Glycosaminoglycan galactosyltransferase. |
| 2.4.1.43 Polygalacturonate 4-alpha-galacturonosyltransferase. | 2.4.1.75 UDP-galacturonosyltransferase. |
| 2.4.1.44 Lipopolysaccharide 3-alpha-galactosyltransferase. | 2.4.1.78 Phosphopolyprenol glucosyltransferase. |
| 2.4.1.45 2-hydroxyacylsphingosine 1-beta-galactosyltransferase. | 2.4.1.79 Galactosylgalactosylglucosylceramide beta-D- acetylgalactosaminyltransferase. |
| 2.4.1.46 1,2-diacylglycerol 3-beta-galactosyltransferase. | 2.4.1.80 Ceramide glucosyltransferase. |
| 2.4.1.47 N-acylsphingosine galactosyltransferase. | 2.4.1.81 Flavone 7-O-beta-glucosyltransferase. |
| 2.4.1.48 Heteroglycan alpha-mannosyltransferase. | 2.4.1.82 Galactinol--sucrose galactosyltransferase. |
| 2.4.1.49 Cellodextrin phosphorylase. | 2.4.1.83 Dolichyl-phosphate beta-D-mannosyltransferase. |
| 2.4.1.50 Procollagen galactosyltransferase. | 2.4.1.85 Cyanohydrin beta-glucosyltransferase. |
| 2.4.1.52 Poly(glycerol-phosphate) alpha-glucosyltransferase. | 2.4.1.86 Glucosaminylgalactosylglucosylceramide beta-galactosyltransferase. |
| 2.4.1.53 Poly(ribitol-phosphate) beta-glucosyltransferase. | 2.4.1.87 N-acetyllactosaminide 3-alpha-galactosyltransferase. |
| 2.4.1.54 Undecaprenyl-phosphate mannosyltransferase. | 2.4.1.88 Globoside alpha-N-acetylgalactosaminyltransferase. |
| 2.4.1.56 Lipopolysaccharide N-acetylglucosaminyltransferase. | 2.4.1.90 N-acetyllactosamine synthase. |
| 2.4.1.57 Phosphatidylinositol alpha-mannosyltransferase. | 2.4.1.91 Flavonol 3-O-glucosyltransferase. |
| 2.4.1.58 Lipopolysaccharide glucosyltransferase I. | 2.4.1.92 (N-acetylneuraminyl)-galactosylglucosylceramide N-acetylgalactosaminyltransferase. |
| 2.4.1.60 Abequosyltransferase. | 2.4.1.94 Protein N-acetylglucosaminyltransferase. |
| 2.4.1.62 Ganglioside galactosyltransferase. | 2.4.1.95 Bilirubin-glucuronoside glucuronosyltransferase. |
| 2.4.1.63 Linamarin synthase. | 2.4.1.96 Sn-glycerol-3-phosphate 1-galactosyltransferase. |
| 2.4.1.64 Alpha,alpha-trehalose phosphorylase. | 2.4.1.97 1,3-beta-D-glucan phosphorylase. |
| 2.4.1.65 3-galactosyl-N-acetylglucosaminide 4-alpha-L-fucosyltransferase. | 2.4.1.99 Sucrose:sucrose fructosyltransferase. |
| 2.4.1.66 Procollagen glucosyltransferase. | 2.4.1.100 2,1-fructan:2,1-fructan 1-fructosyltransferase. |
| 2.4.1.67 Galactinol--raffinose galactosyltransferase. | 2.4.1.101 Alpha-1,3-mannosyl-glycoprotein 2-beta-N- acetylglucosaminyltransferase. |
| 2.4.1.68 Glycoprotein 6-alpha-L-fucosyltransferase. | 2.4.1.102 Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase. |

| | |
|---|---|
| 2.4.1.69 Galactoside 2-alpha-L-fucosyltransferase. | 2.4.1.103 Alizarin 2-beta-glucosyltransferase. |
| 2.4.1.70 Poly(ribitol-phosphate) N-acetylglucosaminyltransferase. | 2.4.1.104 O-dihydroxycoumarin 7-O-glucosyltransferase. |

| | |
|---|---|
| 2.4.1.105 Vitexin beta-glucosyltransferase. | 2.4.1.135 Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase. |
| 2.4.1.106 Isovitexin beta-glucosyltransferase. | 2.4.1.136 Gallate 1-beta-glucosyltransferase. |
| 2.4.1.109 Dolichyl-phosphate-mannose--protein mannosyltransferase. | 2.4.1.137 Sn-glycerol-3-phosphate 2-alpha-galactosyltransferase. |
| 2.4.1.110 tRNA-queuosine beta-mannosyltransferase. | 2.4.1.138 Mannotetraose 2-alpha-N-acetylglucosaminyltransferase. |
| 2.4.1.111 Coniferyl-alcohol glucosyltransferase. | 2.4.1.139 Maltose synthase. |
| 2.4.1.112 Alpha-1,4-glucan-protein synthase (UDP-forming). | 2.4.1.140 Alternansucrase. |
| 2.4.1.113 Alpha-1,4-glucan-protein synthase (ADP-forming). | 2.4.1.141 N-acetylglucosaminyldiphosphodolichol N-acetylglucosaminyltransferase. |
| 2.4.1.114 2-coumarate O-beta-glucosyltransferase. | 2.4.1.142 Chitobiosyldiphosphodolichol beta-mannosyltransferase. |
| 2.4.1.115 Anthocyanidin 3-O-glucosyltransferase. | 2.4.1.143 Alpha-1,6-mannosyl-glycoprotein 2-beta-N- acetylglucosaminyltransferase. |
| 2.4.1.116 Cyanidin-3-rhamnosylglucoside 5-O-glucosyltransferase. | 2.4.1.144 Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase. |
| 2.4.1.117 Dolichyl-phosphate beta-glucosyltransferase. | 2.4.1.145 Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase. |
| 2.4.1.118 Cytokinin 7-beta-glucosyltransferase. | 2.4.1.146 Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase. |
| 2.4.1.119 Dolichyl-diphosphooligosaccharide--protein glycotransferase. | 2.4.1.147 Acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase. |
| 2.4.1.120 Sinapate 1-glucosyltransferase. | 2.4.1.148 Acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase. |
| 2.4.1.121 Indole-3-acetate beta-glucosyltransferase. | 2.4.1.149 N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase. |
| 2.4.1.122 Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase. | 2.4.1.150 N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase. |
| 2.4.1.123 Inositol 3-alpha-galactosyltransferase. | 2.4.1.152 4-galactosyl-N-acetylglucosaminide 3-alpha-L-fucosyltransferase. |
| 2.4.1.125 Sucrose--1,6-alpha-glucan 3(6)-alpha-glucosyltransferase. | 2.4.1.153 Dolichyl-phosphate alpha-N-acetylglucosaminyltransferase. |
| 2.4.1.126 Hydroxycinnamate 4-beta-glucosyltransferase. | 2.4.1.154 Globotriosylceramide beta-1,6-N-acetylgalactosaminyl-transferase. |
| 2.4.1.127 Monoterpenol beta-glucosyltransferase. | 2.4.1.155 Alpha-1,6-mannosyl-glycoprotein 6-beta-N-acetylglucosaminyltransferase. |
| 2.4.1.128 Scopoletin glucosyltransferase. | 2.4.1.156 Indolylacetyl-myo-inositol galactosyltransferase. |
| 2.4.1.129 Peptidoglycan glycosyltransferase. | 2.4.1.157 1,2-diacylglycerol 3-glucosyltransferase. |
| 2.4.1.130 Dolichyl-phosphate-mannose--glycolipid alpha-mannosyltransferase. | 2.4.1.158 13-hydroxydocosanoate 13-beta-glucosyltransferase. |
| 2.4.1.131 Glycolipid 2-alpha-mannosyltransferase. | 2.4.1.159 Flavonol-3-O-glucoside L-rhamnosyltransferase. |
| 2.4.1.132 Glycolipid 3-alpha-mannosyltransferase. | 2.4.1.160 Pyridoxine 5'-O-beta-D-glucosyltransferase. |

| | |
|---|---|
| 2.4.1.133 Xylosylprotein 4-beta-galactosyltransferase. | 2.4.1.161 Oligosaccharide 4-alpha-D-glucosyltransferase. |
| 2.4.1.134 Galactosylxylosylprotein 3-beta-galactosyltransferase. | 2.4.1.162 Aldose beta-D-fructosyltransferase. |

| | |
|---|---|
| 2.4.1.163 Beta-galactosyl-N-acetylglucosaminylgalactosylglucosyl-ceramide beta-1,3-acetylglucosaminyltransferase. | 2.4.1.191 Luteolin-7-O-diglucuronide 4'-O-glucuronosyltransferase. |
| 2.4.1.164 Galactosyl-N-acetylglucosaminylgalactosylglucosyl-ceramide beta-1,6-N-acetylglucosaminyltransferase. | 2.4.1.192 Nuatigenin 3-beta-glucosyltransferase. |
| 2.4.1.165 N-acetylneuraminylgalactosylglucosylceramide beta-1,4-N-acetylgalactosaminyltransferase. | 2.4.1.193 Sarsapogenin 3-beta-glucosyltransferase. |
| 2.4.1.166 Raffinose--raffinose alpha-galactosyltransferase. | 2.4.1.194 4-hydroxybenzoate 4-O-beta-D-glucosyltransferase. |
| 2.4.1.167 Sucrose 6(F)-alpha-galactosyltransferase. | 2.4.1.195 Thiohydroximate beta-D-glucosyltransferase. |
| 2.4.1.168 Xyloglucan 4-glucosyltransferase. | 2.4.1.196 Nicotinate glucosyltransferase. |
| 2.4.1.170 Isoflavone 7-O-glucosyltransferase. | 2.4.1.197 High-mannose-oligosaccharide beta-1,4-N-acetylglucosaminyltransferase. |
| 2.4.1.171 Methyl-ONN-azoxymethanol beta-D-glucosyltransferase. | 2.4.1.198 Phosphatidylinositol N-acetylglucosaminyltransferase. |
| 2.4.1.172 Salicyl-alcohol beta-D-glucosyltransferase. | 2.4.1.199 Beta-mannosylphosphodecaprenol—mannooligosaccharide 6-mannosyltransferase. |
| 2.4.1.173 Sterol 3-beta-glucosyltransferase. | 2.4.1.201 Alpha-1,6-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase. |
| 2.4.1.174 Glucuronylgalactosylproteoglycan 4-beta-N-acetylgalactosaminyltransferase. | 2.4.1.202 2,4-dihydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one 2-D-glucosyltransferase. |
| 2.4.1.175 Glucuronosyl-N-acetylgalactosaminyl-proteoglycan 4-beta-N-acetylgalactosaminyltransferase. | 2.4.1.203 Trans-zeatin O-beta-D-glucosyltransferase. |
| 2.4.1.176 Gibberellin beta-D-glucosyltransferase. | 2.4.1.205 Galactogen 6-beta-galactosyltransferase. |
| 2.4.1.177 Cinnamate beta-D-glucosyltransferase. | 2.4.1.206 Lactosylceramide 1,3-N-acetyl-beta-D-glucosaminyltransferase. |
| 2.4.1.178 Hydroxymandelonitrile glucosyltransferase. | 2.4.1.207 Xyloglucan:xyloglucosyl transferase. |
| 2.4.1.179 Lactosylceramide beta-1,3-galactosyltransferase. | 2.4.1.208 Diglucosyl diacylglycerol synthase. |
| 2.4.1.180 Lipopolysaccharide N-acetylmannosaminouronosyltransferase. | 2.4.1.209 Cis-p-coumarate glucosyltransferase. |
| 2.4.1.181 Hydroxyanthraquinone glucosyltransferase. | 2.4.1.210 Limonoid glucosyltransferase. |
| 2.4.1.182 Lipid-A-disaccharide synthase. | 2.4.1.211 1,3-beta-galactosyl-N-acetylhexosamine phosphorylase. |
| 2.4.1.183 Alpha-1,3-glucan synthase. | 2.4.1.212 Hyaluronan synthase. |
| 2.4.1.184 Galactolipid galactosyltransferase. | 2.4.1.213 Glucosylglycerol-phosphate synthase. |
| 2.4.1.185 Flavanone 7-O-beta-glucosyltransferase. | 2.4.1.214 Glycoprotein 3-alpha-L-fucosyltransferase. |
| 2.4.1.186 Glycogenin glucosyltransferase. | 2.4.1.215 Cis-zeatin O-beta-D-glucosyltransferase. |

| | |
|---|---|
| 2.4.1.187 N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase. | 2.4.1.216 Trehalose 6-phosphate phosphorylase. |
| 2.4.1.188 N-acetylglucosaminyldiphosphoundecaprenol glucosyltransferase. | 2.4.1.217 Mannosyl-3-phosphoglycerate synthase. |
| 2.4.1.189 Luteolin 7-O-glucuronosyltransferase. | 2.4.1.218 Hydroquinone glucosyltransferase. |
| 2.4.1.190 Luteolin-7-O-glucuronide 7-O-glucuronosyltransferase. | 2.4.1.219 Vomilenine glucosyltransferase. |

| | |
|---|---|
| 2.4.1.220 Indoxyl-UDPG glucosyltransferase. | 2.4.2.19 Nicotinate-nucleotide diphosphorylase (carboxylating). |
| 2.4.1.221 Peptide-O-fucosyltransferase. | 2.4.2.20 Dioxotetrahydropyrimidine phosphoribosyltransferase. |
| 2.4.1.222 O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase. | 2.4.2.21 Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase. |
| 2.4.1.223 Glucuronyl-galactosyl-proteoglycan 4-alpha-N-acetylglucosaminyltransferase. | 2.4.2.22 Xanthine phosphoribosyltransferase. |
| 2.4.1.224 Glucuronosyl-N-acetylglucosaminyl-proteoglycan 4-alpha-N-acetylglucosaminyltransferase. | 2.4.2.23 Deoxyuridine phosphorylase. |
| 2.4.1.225 N-acetylglucosaminyl-proteoglycan 4-beta-glucuronosyltransferase. | 2.4.2.24 1,4-beta-D-xylan synthase. |
| 2.4.1.226 N-acetylgalactosaminyl-proteoglycan 3-beta-glucuronosyltransferase. | 2.4.2.25 Flavone apiosyltransferase. |
| 2.4.1.227 Undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase. | 2.4.2.26 Protein xylosyltransferase. |
| 2.4.1.228 Lactosylceramide 4-alpha-galactosyltransferase. | 2.4.2.27 dTDP-dihydrostreptose--streptidine-6-phosphate dihydrostreptosyltransferase. |
| 2.4.1.229 [Skp1-protein]-hydroxyproline N-acetylglucosaminyltransferase. | 2.4.2.28 S-methyl-5-thioadenosine phosphorylase. |
| 2.4.1.230 Kojibiose phosphorylase. | 2.4.2.29 Queuine tRNA-ribosyltransferase. |
| 2.4.1.231 Alpha,alpha-trehalose phosphorylase (configuration-retaining). | 2.4.2.30 NAD(+) ADP-ribosyltransferase. |
| 2.4.1.232 Initiation-specific alpha-1,6-mannosyltransferase. | 2.4.2.31 NAD(P)(+)--arginine ADP-ribosyltransferase. |
| 2.4.2.1 Purine-nucleoside phosphorylase. | 2.4.2.32 Dolichyl-phosphate D-xylosyltransferase. |
| 2.4.2.2 Pyrimidine-nucleoside phosphorylase. | 2.4.2.33 Dolichyl-xylosyl-phosphate--protein xylosyltransferase. |
| 2.4.2.3 Uridine phosphorylase. | 2.4.2.34 Indolylacetylinositol arabinosyltransferase. |
| 2.4.2.4 Thymidine phosphorylase. | 2.4.2.35 Flavonol-3-O-glycoside xylosyltransferase. |
| 2.4.2.5 Nucleoside ribosyltransferase. | 2.4.2.36 NAD(+)--diphthamide ADP-ribosyltransferase. |
| 2.4.2.6 Nucleoside deoxyribosyltransferase. | 2.4.2.37 NAD(+)--dinitrogen-reductase ADP-D-ribosyltransferase. |
| 2.4.2.7 Adenine phosphoribosyltransferase. | 2.4.2.38 Glycoprotein 2-beta-D-xylosyltransferase. |
| 2.4.2.8 Hypoxanthine phosphoribosyltransferase. | 2.4.2.39 Xyloglucan 6-xylosyltransferase. |
| 2.4.2.9 Uracil phosphoribosyltransferase. | 2.4.2.40 Zeatin O-beta-D-xylosyltransferase. |
| 2.4.2.10 Orotate phosphoribosyltransferase. | 2.4.99.1 Beta-galactoside alpha-2,6-sialyltransferase. |
| 2.4.2.11 Nicotinate phosphoribosyltransferase. | 2.4.99.2 Monosialoganglioside sialyltransferase. |
| 2.4.2.12 Nicotinamide phosphoribosyltransferase. | 2.4.99.3 Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase. |

| | |
|---|---|
| 2.4.2.14 Amidophosphoribosyltransferase. | 2.4.99.4 Beta-galactoside alpha-2,3-sialyltransferase. |
| 2.4.2.15 Guanosine phosphorylase. | 2.4.99.5 Galactosyldiacylglycerol alpha-2,3-sialyltransferase. |
| 2.4.2.16 Urate-ribonucleotide phosphorylase. | 2.4.99.6 N-acetyllactosaminide alpha-2,3-sialyltransferase. |
| 2.4.2.17 ATP phosphoribosyltransferase. | 2.4.99.7 (Alpha-N-acetylneuraminyl-2,3-beta-galactosyl-1,3)-N-acetyl-galactosaminide 6-alpha-sialyltransferase. |
| 2.4.2.18 Anthranilate phosphoribosyltransferase. | 2.4.99.8 Alpha-N-acetylneuraminate alpha-2,8-sialyltransferase. |

| | |
|---|---|
| 2.4.99.9 Lactosylceramide alpha-2,3-sialyltransferase. | 2.5.1.42 Geranylgeranylglycerol-phosphate geranylgeranyltransferase. |
| 2.4.99.10 Neolactotetraosylceramide alpha-2,3-sialyltransferase. | 2.5.1.43 Nicotianamine synthase. |
| 2.4.99.11 Lactosylceramide alpha-2,6-N-sialyltransferase. | 2.5.1.44 Homospermidine synthase. |
| 2.5.1.1 Dimethylallyltranstransferase. | 2.5.1.45 Homospermidine synthase (spermidine-specific). |
| 2.5.1.2 Thiamine pyridinylase. | 2.5.1.46 Deoxyhypusine synthase. |
| 2.5.1.3 Thiamine-phosphate diphosphorylase. | 2.5.1.47 Cysteine synthase. |
| 2.5.1.4 Adenosylmethionine cyclotransferase. | 2.5.1.48 Cystathionine gamma-synthase. |
| 2.5.1.5 Galactose-6-sulfurylase. | 2.5.1.49 O-acetylhomoserine aminocarboxypropyltransferase. |
| 2.5.1.6 Methionine adenosyltransferase. | 2.5.1.50 Zeatin 9-aminocarboxyethyltransferase. |
| 2.5.1.7 UDP-N-acetylglucosamine 1-carboxyvinyltransferase. | 2.5.1.51 Beta-pyrazolylalanine synthase. |
| 2.5.1.8 tRNA isopentenyltransferase. | 2.5.1.52 L-mimosine synthase. |
| 2.5.1.9 Riboflavin synthase. | 2.5.1.53 Uracilylalanine synthase. |
| 2.5.1.10 Geranyltranstransferase. | 2.5.1.54 3-deoxy-7-phosphoheptulonate synthase. |
| 2.5.1.11 Trans-octaprenyltranstransferase. | 2.5.1.55 3-deoxy-8-phosphooctulonate synthase. |
| 2.5.1.15 Dihydropteroate synthase. | 2.5.1.56 N-acetylneuraminate synthase. |
| 2.5.1.16 Spermidine synthase. | 2.5.1.57 N-acylneuraminate-9-phosphate synthase. |
| 2.5.1.17 Cob(I)yrinic acid a,c-diamide adenosyltransferase. | 2.5.1.58 Protein farnesyltransferase. |
| 2.5.1.18 Glutathione transferase. | 2.5.1.59 Protein geranylgeranyltransferase type I. |
| 2.5.1.19 3-phosphoshikimate 1-carboxyvinyltransferase. | 2.5.1.60 Protein geranylgeranyltransferase type II. |
| 2.5.1.20 Rubber cis-polyprenylcistransferase. | 2.5.1.61 Hydroxymethylbilane synthase. |
| 2.5.1.21 Farnesyl-diphosphate farnesyltransferase. | 2.5.1.62 Chlorophyll synthase. |
| 2.5.1.22 Spermine synthase. | 2.5.1.63 Adenosyl-fluoride synthase. |
| 2.5.1.23 Sym-norspermidine synthase. | 2.5.1.64 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase. |
| 2.5.1.24 Discadenine synthase. | 2.6.1.1 Aspartate transaminase. |
| 2.5.1.25 tRNA-uridine aminocarboxypropyltransferase. | 2.6.1.2 Alanine transaminase. |
| 2.5.1.26 Alkylglycerone-phosphate synthase. | 2.6.1.3 Cysteine transaminase. |
| 2.5.1.27 Adenylate dimethylallyltransferase. | 2.6.1.4 Glycine transaminase. |
| 2.5.1.28 Dimethylallylcistransferase. | 2.6.1.5 Tyrosine transaminase. |
| 2.5.1.29 Farnesyltranstransferase. | 2.6.1.6 Leucine transaminase. |
| 2.5.1.30 Trans-hexaprenyltranstransferase. | 2.6.1.7 Kynurenine--oxoglutarate transaminase. |
| 2.5.1.31 Di-trans,poly-cis-decaprenylcistransferase. | 2.6.1.8 2,5-diaminovalerate transaminase. |

| | |
|---|---|
| 2.5.1.32 Geranylgeranyl-diphosphate geranylgeranyltransferase. | 2.6.1.9 Histidinol-phosphate transaminase. |
| 2.5.1.33 Trans-pentaprenyltranstransferase. | 2.6.1.11 Acetylornithine transaminase. |
| 2.5.1.34 Tryptophan dimethylallyltransferase. | 2.6.1.12 Alanine--oxo-acid transaminase. |
| 2.5.1.35 Aspulvinone dimethylallyltransferase. | 2.6.1.13 Ornithine--oxo-acid transaminase. |
| 2.5.1.36 Trihydroxypterocarpan dimethylallyltransferase. | 2.6.1.14 Asparagine--oxo-acid transaminase. |
| 2.5.1.38 Isonocardicin synthase. | 2.6.1.15 Glutamine--pyruvate transaminase. |
| 2.5.1.39 4-hydroxybenzoate nonaprenyltransferase. | 2.6.1.16 Glutamine--fructose-6-phosphate transaminase (isomerizing). |
| 2.5.1.41 Phosphoglycerol geranylgeranyltransferase. | 2.6.1.17 Succinyldiaminopimelate transaminase. |

| | |
|---|---|
| 2.6.1.18 Beta-alanine--pyruvate transaminase. | 2.6.1.62 Adenosylmethionine--8-amino-7-oxononanoate transaminase. |
| 2.6.1.19 4-aminobutyrate transaminase. | 2.6.1.63 Kynurenine--glyoxylate transaminase. |
| 2.6.1.21 D-alanine transaminase. | 2.6.1.64 Glutamine--phenylpyruvate transaminase. |
| 2.6.1.22 (S)-3-amino-2-methylpropionate transaminase. | 2.6.1.65 N(6)-acetyl-beta-lysine transaminase. |
| 2.6.1.23 4-hydroxyglutamate transaminase. | 2.6.1.66 Valine--pyruvate transaminase. |
| 2.6.1.24 Diiodotyrosine transaminase. | 2.6.1.67 2-aminohexanoate transaminase. |
| 2.6.1.26 Thyroid-hormone transaminase. | 2.6.1.68 Ornithine(lysine) transaminase. |
| 2.6.1.27 Tryptophan transaminase. | 2.6.1.70 Aspartate--phenylpyruvate transaminase. |
| 2.6.1.28 Tryptophan--phenylpyruvate transaminase. | 2.6.1.71 Lysine--pyruvate 6-transaminase. |
| 2.6.1.29 Diamine transaminase. | 2.6.1.72 D-4-hydroxyphenylglycine transaminase. |
| 2.6.1.30 Pyridoxamine--pyruvate transaminase. | 2.6.1.73 Methionine--glyoxylate transaminase. |
| 2.6.1.31 Pyridoxamine--oxaloacetate transaminase. | 2.6.1.74 Cephalosporin-C transaminase. |
| 2.6.1.32 Valine--3-methyl-2-oxovalerate transaminase. | 2.6.1.75 Cysteine-conjugate transaminase. |
| 2.6.1.33 dTDP-4-amino-4,6-dideoxy-D-glucose transaminase. | 2.6.1.76 Diaminobutyrate--2-oxoglutarate transaminase. |
| 2.6.1.34 UDP-2-acetamido-4-amino-2,4,6-trideoxyglucose transaminase. | 2.6.1.77 Taurine--pyruvate aminotransferase. |
| 2.6.1.35 Glycine--oxaloacetate transaminase. | 2.6.3.1 Oximinotransferase. |
| 2.6.1.36 L-lysine 6-transaminase. | 2.6.99.1 dATP(dGTP)--DNA purinetransferase. |
| 2.6.1.37 2-aminoethylphosphonate--pyruvate transaminase. | 2.7.1.1 Hexokinase. |
| 2.6.1.38 Histidine transaminase. | 2.7.1.2 Glucokinase. |
| 2.6.1.39 2-aminoadipate transaminase. | 2.7.1.3 Ketohexokinase. |
| 2.6.1.40 (R)-3-amino-2-methylpropionate--pyruvate transaminase. | 2.7.1.4 Fructokinase. |
| 2.6.1.41 D-methionine--pyruvate transaminase. | 2.7.1.5 Rhamnulokinase. |
| 2.6.1.42 Branched-chain-amino-acid transaminase. | 2.7.1.6 Galactokinase. |
| 2.6.1.43 Aminolevulinate transaminase. | 2.7.1.7 Mannokinase. |
| 2.6.1.44 Alanine--glyoxylate transaminase. | 2.7.1.8 Glucosamine kinase. |
| 2.6.1.45 Serine--glyoxylate transaminase. | 2.7.1.10 Phosphoglucokinase. |
| 2.6.1.46 Diaminobutyrate--pyruvate transaminase. | 2.7.1.11 6-phosphofructokinase. |
| 2.6.1.47 Alanine--oxomalonate transaminase. | 2.7.1.12 Gluconokinase. |
| 2.6.1.48 5-aminovalerate transaminase. | 2.7.1.13 Dehydrogluconokinase. |
| 2.6.1.49 Dihydroxyphenylalanine transaminase. | 2.7.1.14 Sedoheptulokinase. |

| | |
|---|---|
| 2.6.1.50 Glutamine--scyllo-inositol transaminase. | 2.7.1.15 Ribokinase. |
| 2.6.1.51 Serine--pyruvate transaminase. | 2.7.1.16 Ribulokinase. |
| 2.6.1.52 Phosphoserine transaminase. | 2.7.1.17 Xylulokinase. |
| 2.6.1.54 Pyridoxamine-phosphate transaminase. | 2.7.1.18 Phosphoribokinase. |
| 2.6.1.55 Taurine-2-oxoglutarate transaminase. | 2.7.1.19 Phosphoribulokinase. |
| 2.6.1.56 1D-1-guanidino-3-amino-1,3-dideoxy-scyllo-inositol transaminase. | 2.7.1.20 Adenosine kinase. |
| 2.6.1.57 Aromatic-amino-acid transaminase. | 2.7.1.21 Thymidine kinase. |
| 2.6.1.58 Phenylalanine(histidine) transaminase. | 2.7.1.22 Ribosylnicotinamide kinase. |
| 2.6.1.59 dTDP-4-amino-4,6-dideoxygalactose transaminase. | 2.7.1.23 NAD(+) kinase. |
| 2.6.1.60 Aromatic-amino-acid--glyoxylate transaminase. | 2.7.1.24 Dephospho-CoA kinase. |

| | |
|---|---|
| 2.7.1.25 Adenylyl-sulfate kinase. | 2.7.1.69 Protein-N(pi)-phosphohistidine--sugar phosphotransferase. |
| 2.7.1.26 Riboflavin kinase. | 2.7.1.71 Shikimate kinase. |
| 2.7.1.27 Erythritol kinase. | 2.7.1.72 Streptomycin 6-kinase. |
| 2.7.1.28 Triokinase. | 2.7.1.73 Inosine kinase. |
| 2.7.1.29 Glycerone kinase. | 2.7.1.74 Deoxycytidine kinase. |
| 2.7.1.30 Glycerol kinase. | 2.7.1.76 Deoxyadenosine kinase. |
| 2.7.1.31 Glycerate kinase. | 2.7.1.77 Nucleoside phosphotransferase. |
| 2.7.1.32 Choline kinase. | 2.7.1.78 Polynucleotide 5'-hydroxy-kinase. |
| 2.7.1.33 Pantothenate kinase. | 2.7.1.79 Diphosphate--glycerol phosphotransferase. |
| 2.7.1.34 Pantetheine kinase. | 2.7.1.80 Diphosphate--serine phosphotransferase. |
| 2.7.1.35 Pyridoxal kinase. | 2.7.1.81 Hydroxylysine kinase. |
| 2.7.1.36 Mevalonate kinase. | 2.7.1.82 Ethanolamine kinase. |
| 2.7.1.37 Protein kinase. | 2.7.1.83 Pseudouridine kinase. |
| 2.7.1.38 Phosphorylase kinase. | 2.7.1.84 Alkylglycerone kinase. |
| 2.7.1.39 Homoserine kinase. | 2.7.1.85 Beta-glucoside kinase. |
| 2.7.1.40 Pyruvate kinase. | 2.7.1.86 NADH kinase. |
| 2.7.1.41 Glucose-1-phosphate phosphodismutase. | 2.7.1.87 Streptomycin 3"-kinase. |
| 2.7.1.42 Riboflavin phosphotransferase. | 2.7.1.88 Dihydrostreptomycin-6-phosphate 3'-alpha-kinase. |
| 2.7.1.43 Glucuronokinase. | 2.7.1.89 Thiamine kinase. |
| 2.7.1.44 Galacturonokinase. | 2.7.1.90 Diphosphate--fructose-6-phosphate 1-phosphotransferase. |
| 2.7.1.45 2-dehydro-3-deoxygluconokinase. | 2.7.1.91 Sphinganine kinase. |
| 2.7.1.46 L-arabinokinase. | 2.7.1.92 5-dehydro-2-deoxygluconokinase. |
| 2.7.1.47 D-ribulokinase. | 2.7.1.93 Alkylglycerol kinase. |
| 2.7.1.48 Uridine kinase. | 2.7.1.94 Acylglycerol kinase. |
| 2.7.1.49 Hydroxymethylpyrimidine kinase. | 2.7.1.95 Kanamycin kinase. |
| 2.7.1.50 Hydroxyethylthiazole kinase. | 2.7.1.99 [Pyruvate dehydrogenase (lipoamide)] kinase. |
| 2.7.1.51 L-fuculokinase. | 2.7.1.100 S-methyl-5-thioribose kinase. |
| 2.7.1.52 Fucokinase. | 2.7.1.101 Tagatose kinase. |
| 2.7.1.53 L-xylulokinase. | 2.7.1.102 Hamamelose kinase. |
| 2.7.1.54 D-arabinokinase. | 2.7.1.103 Viomycin kinase. |
| 2.7.1.55 Allose kinase. | 2.7.1.104 Diphosphate--protein phosphotransferase. |
| 2.7.1.56 1-phosphofructokinase. | 2.7.1.105 6-phosphofructo-2-kinase. |
| 2.7.1.58 2-dehydro-3-deoxygalactonokinase. | 2.7.1.106 Glucose-1,6-bisphosphate synthase. |
| 2.7.1.59 N-acetylglucosamine kinase. | 2.7.1.107 Diacylglycerol kinase. |
| 2.7.1.60 N-acylmannosamine kinase. | 2.7.1.108 Dolichol kinase. |

| | |
|---|---|
| 2.7.1.61 Acyl-phosphate--hexose phosphotransferase. | 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase (NADPH)] kinase. |
| 2.7.1.62 Phosphoramidate--hexose phosphotransferase. | 2.7.1.110 Dephospho-[reductase kinase] kinase. |
| 2.7.1.63 Polyphosphate--glucose phosphotransferase. | 2.7.1.112 Protein-tyrosine kinase. |
| 2.7.1.64 Inositol 3-kinase. | 2.7.1.113 Deoxyguanosine kinase. |
| 2.7.1.65 Scyllo-inosamine 4-kinase. | 2.7.1.114 AMP--thymidine kinase. |
| 2.7.1.66 Undecaprenol kinase. | 2.7.1.115 [3-methyl-2-oxobutanoate dehydrogenase (lipoamide)] kinase. |
| 2.7.1.67 1-phosphatidylinositol 4-kinase. | 2.7.1.116 [Isocitrate dehydrogenase (NADP+)] kinase. |
| 2.7.1.68 1-phosphatidylinositol-4-phosphate 5-kinase. | 2.7.1.117 [Myosin light-chain] kinase. |

| | |
|---|---|
| 2.7.1.118 ADP--thymidine kinase. | 2.7.2.6 Formate kinase. |
| 2.7.1.119 Hygromycin-B kinase. | 2.7.2.7 Butyrate kinase. |
| 2.7.1.120 Caldesmon kinase. | 2.7.2.8 Acetylglutamate kinase. |
| 2.7.1.121 Phosphoenolpyruvate--glycerone phosphotransferase. | 2.7.2.10 Phosphoglycerate kinase (GTP). |
| 2.7.1.122 Xylitol kinase. | 2.7.2.11 Glutamate 5-kinase. |
| 2.7.1.123 Calcium/calmodulin-dependent protein kinase. | 2.7.2.12 Acetate kinase (diphosphate). |
| 2.7.1.124 [Tyrosine 3-monooxygenase] kinase. | 2.7.2.13 Glutamate 1-kinase. |
| 2.7.1.125 Rhodopsin kinase. | 2.7.2.14 Branched-chain-fatty-acid kinase. |
| 2.7.1.126 [Beta-adrenergic-receptor] kinase. | 2.7.3.1 Guanidinoacetate kinase. |
| 2.7.1.127 Inositol-trisphosphate 3-kinase. | 2.7.3.2 Creatine kinase. |
| 2.7.1.128 [Acetyl-CoA carboxylase] kinase. | 2.7.3.3 Arginine kinase. |
| 2.7.1.129 [Myosin heavy-chain] kinase. | 2.7.3.4 Taurocyamine kinase. |
| 2.7.1.130 Tetraacyldisaccharide 4'-kinase. | 2.7.3.5 Lombricine kinase. |
| 2.7.1.131 [Low-density lipoprotein receptor] kinase. | 2.7.3.6 Hypotaurocyamine kinase. |
| 2.7.1.132 Tropomyosin kinase. | 2.7.3.7 Opheline kinase. |
| 2.7.1.134 Inositol-tetrakisphosphate 1-kinase. | 2.7.3.8 Ammonia kinase. |
| 2.7.1.135 [Tau protein] kinase. | 2.7.3.9 Phosphoenolpyruvate--protein phosphotransferase. |
| 2.7.1.136 Macrolide 2'-kinase. | 2.7.3.10 Agmatine kinase. |
| 2.7.1.137 Phosphatidylinositol 3-kinase. | 2.7.3.11 Protein-histidine pros-kinase. |
| 2.7.1.138 Ceramide kinase. | 2.7.3.12 Protein-histidine tele-kinase. |
| 2.7.1.140 Inositol-tetrakisphosphate 5-kinase. | 2.7.4.1 Polyphosphate kinase. |
| 2.7.1.141 [RNA-polymerase]-subunit kinase. | 2.7.4.2 Phosphomevalonate kinase. |
| 2.7.1.142 Glycerol-3-phosphate--glucose phosphotransferase. | 2.7.4.3 Adenylate kinase. |
| 2.7.1.143 Diphosphate-purine nucleoside kinase. | 2.7.4.4 Nucleoside-phosphate kinase. |
| 2.7.1.144 Tagatose-6-phosphate kinase. | 2.7.4.6 Nucleoside-diphosphate kinase. |
| 2.7.1.145 Deoxynucleoside kinase. | 2.7.4.7 Phosphomethylpyrimidine kinase. |
| 2.7.1.146 ADP-specific phosphofructokinase. | 2.7.4.8 Guanylate kinase. |
| 2.7.1.147 ADP-specific glucokinase. | 2.7.4.9 dTMP kinase. |
| 2.7.1.148 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase. | 2.7.4.10 Nucleoside-triphosphate--adenylate kinase. |
| 2.7.1.149 1-phosphatidylinositol-5-phosphate 4-kinase. | 2.7.4.11 (Deoxy)adenylate kinase. |
| 2.7.1.150 1-phosphatidylinositol-3-phosphate 5-kinase. | 2.7.4.12 T(2)-induced deoxynucleotide kinase. |

| | |
|---|---|
| 2.7.1.151 Inositol-polyphosphate multikinase. | 2.7.4.13 (Deoxy)nucleoside-phosphate kinase. |
| 2.7.1.153 Phosphatidylinositol-4,5-bisphosphate 3-kinase. | 2.7.4.14 Cytidylate kinase. |
| 2.7.1.154 Phosphatidylinositol-4-phosphate 3-kinase. | 2.7.4.15 Thiamine-diphosphate kinase. |
| 2.7.1.155 Diphosphoinositol-pentakisphosphate kinase. | 2.7.4.16 Thiamine-phosphate kinase. |
| 2.7.1.156 Adenosylcobinamide kinase. | 2.7.4.17 3-phosphoglyceroyl-phosphate--polyphosphate phosphotransferase. |
| 2.7.2.1 Acetate kinase. | 2.7.4.18 Farnesyl-diphosphate kinase. |
| 2.7.2.2 Carbamate kinase. | 2.7.4.19 5-methyldeoxycytidine-5'-phosphate kinase. |
| 2.7.2.3 Phosphoglycerate kinase. | 2.7.4.20 Dolichyl-diphosphate--polyphosphate phosphotransferase. |
| 2.7.2.4 Aspartate kinase. | 2.7.4.21 Inositol-hexakisphosphate kinase. |

| | |
|---|---|
| 2.7.6.1 Ribose-phosphate diphosphokinase. | 2.7.7.34 Glucose-1-phosphate guanylyltransferase. |
| 2.7.6.2 Thiamine diphosphokinase. | 2.7.7.35 Ribose-5-phosphate adenylyltransferase. |
| 2.7.6.3 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase. | 2.7.7.36 Aldose-1-phosphate adenylyltransferase. |
| 2.7.6.4 Nucleotide diphosphokinase. | 2.7.7.37 Aldose-1-phosphate nucleotidyltransferase. |
| 2.7.6.5 GTP diphosphokinase. | 2.7.7.38 3-deoxy-manno-octulosonate cytidylyltransferase. |
| 2.7.7.1 Nicotinamide-nucleotide adenylyltransferase. | 2.7.7.39 Glycerol-3-phosphate cytidylyltransferase. |
| 2.7.7.2 FMN adenylyltransferase. | 2.7.7.40 D-ribitol-5-phosphate cytidylyltransferase. |
| 2.7.7.3 Pantetheine-phosphate adenylyltransferase. | 2.7.7.41 Phosphatidate cytidylyltransferase. |
| 2.7.7.4 Sulfate adenylyltransferase. | 2.7.7.42 [Glutamate--ammonia-ligase] adenylyltransferase. |
| 2.7.7.5 Sulfate adenylyltransferase (ADP). | 2.7.7.43 N-acylneuraminate cytidylyltransferase. |
| 2.7.7.6 DNA-directed RNA polymerase. | 2.7.7.44 Glucuronate-1-phosphate uridylyltransferase. |
| 2.7.7.7 DNA-directed DNA polymerase. | 2.7.7.45 Guanosine-triphosphate guanylyltransferase. |
| 2.7.7.8 Polyribonucleotide nucleotidyltransferase. | 2.7.7.46 Gentamicin 2"-nucleotidyltransferase. |
| 2.7.7.9 UTP--glucose-1-phosphate uridylyltransferase. | 2.7.7.47 Streptomycin 3"-adenylyltransferase. |
| 2.7.7.10 UTP--hexose-1-phosphate uridylyltransferase. | 2.7.7.48 RNA-directed RNA polymerase. |
| 2.7.7.11 UTP--xylose-1-phosphate uridylyltransferase. | 2.7.7.49 RNA-directed DNA polymerase. |
| 2.7.7.12 UDP-glucose--hexose-1-phosphate uridylyltransferase. | 2.7.7.50 mRNA guanylyltransferase. |
| 2.7.7.13 Mannose-1-phosphate guanylyltransferase. | 2.7.7.51 Adenylylsulfate--ammonia adenylyltransferase. |
| 2.7.7.14 Ethanolamine-phosphate cytidylyltransferase. | 2.7.7.52 RNA uridylyltransferase. |
| 2.7.7.15 Choline-phosphate cytidylyltransferase. | 2.7.7.53 ATP adenylyltransferase. |
| 2.7.7.18 Nicotinate-nucleotide adenylyltransferase. | 2.7.7.54 Phenylalanine adenylyltransferase. |
| 2.7.7.19 Polynucleotide adenylyltransferase. | 2.7.7.55 Anthranilate adenylyltransferase. |
| 2.7.7.21 tRNA cytidylyltransferase. | 2.7.7.56 tRNA nucleotidyltransferase. |

| | |
|---|---|
| 2.7.7.22 Mannose-1-phosphate guanylyltransferase (GDP). | 2.7.7.57 N-methylphosphoethanolamine cytidylyltransferase. |
| 2.7.7.23 UDP-N-acetylglucosamine diphosphorylase. | 2.7.7.58 (2,3-dihydroxybenzoyl)adenylate synthase. |
| 2.7.7.24 Glucose-1-phosphate thymidylyltransferase. | 2.7.7.59 [Protein-PII] uridylyltransferase. |
| 2.7.7.25 tRNA adenylyltransferase. | 2.7.7.60 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase. |
| 2.7.7.27 Glucose-1-phosphate adenylyltransferase. | 2.7.7.61 Holo-ACP synthase. |
| 2.7.7.28 Nucleoside-triphosphate-aldose 1-phosphate nucleotidyltransferase. | 2.7.7.62 Adenosylcobinamide-phosphate guanylyltransferase. |
| 2.7.7.30 Fucose-1-phosphate guanylyltransferase. | 2.7.8.1 Ethanolaminephosphotransferase. |
| 2.7.7.31 DNA nucleotidylexotransferase. | 2.7.8.2 Diacylglycerol cholinephosphotransferase. |
| 2.7.7.32 Galactose-1-phosphate thymidylyltransferase. | 2.7.8.3 Ceramide cholinephosphotransferase. |
| 2.7.7.33 Glucose-1-phosphate cytidylyltransferase. | 2.7.8.4 Serine-phosphoethanolamine synthase. |

| | |
|---|---|
| 2.7.8.5 CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase. | 2.8.2.1 Aryl sulfotransferase. |
| 2.7.8.6 Undecaprenyl-phosphate galactose phosphotransferase. | 2.8.2.2 Alcohol sulfotransferase. |
| 2.7.8.7 Holo-[acyl-carrier-protein] synthase. | 2.8.2.3 Amine sulfotransferase. |
| 2.7.8.8 CDP-diacylglycerol--serine O-phosphatidyltransferase. | 2.8.2.4 Estrone sulfotransferase. |
| 2.7.8.9 Phosphomannan mannosephosphotransferase. | 2.8.2.5 Chondroitin 4-sulfotransferase. |
| 2.7.8.10 Sphingosine cholinephosphotransferase. | 2.8.2.6 Choline sulfotransferase. |
| 2.7.8.11 CDP-diacylglycerol--inositol 3-phosphatidyltransferase. | 2.8.2.7 UDP-N-acetylgalactosamine-4-sulfate sulfotransferase. |
| 2.7.8.12 CDP-glycerol glycerophosphotransferase. | 2.8.2.8 [Heparan sulfate]-glucosamine N-sulfotransferase. |
| 2.7.8.13 Phospho-N-acetylmuramoyl-pentapeptide-transferase. | 2.8.2.9 Tyrosine-ester sulfotransferase. |
| 2.7.8.14 CDP-ribitol ribitolphosphotransferase. | 2.8.2.10 Renilla-luciferin sulfotransferase. |
| 2.7.8.15 UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase. | 2.8.2.11 Galactosylceramide sulfotransferase. |
| 2.7.8.17 UDP-N-acetylglucosamine--lysosomal-enzyme N-acetylglucosaminephosphotransferase. | 2.8.2.13 Psychosine sulfotransferase. |
| 2.7.8.18 UDP-galactose--UDP-N-acetylglucosamine galactose phosphotransferase. | 2.8.2.14 Bile-salt sulfotransferase. |
| 2.7.8.19 UDP-glucose--glycoprotein glucose phosphotransferase. | 2.8.2.15 Steroid sulfotransferase. |
| 2.7.8.20 Phosphatidylglycerol--membrane-oligosaccharide glycerophosphotransferase. | 2.8.2.16 Thiol sulfotransferase. |
| 2.7.8.21 Membrane-oligosaccharide glycerophosphotransferase. | 2.8.2.17 Chondroitin 6-sulfotransferase. |
| 2.7.8.22 1-alkenyl-2-acylglycerol choline phosphotransferase. | 2.8.2.18 Cortisol sulfotransferase. |
| 2.7.8.23 Carboxyvinyl-carboxyphosphonate phosphorylmutase. | 2.8.2.19 Triglucosylalkylacylglycerol sulfotransferase. |
| 2.7.8.24 Phosphatidylcholine synthase. | 2.8.2.20 Protein-tyrosine sulfotransferase. |

| | |
|---|---|
| 2.7.8.25 Triphosphoribosyl-dephospho-CoA synthase. | 2.8.2.21 Keratan sulfotransferase. |
| 2.7.8.26 Adenosylcobinamide-GDP ribazoletransferase. | 2.8.2.22 Arylsulfate sulfotransferase. |
| 2.7.9.1 Pyruvate, phosphate dikinase. | 2.8.2.23 [Heparan sulfate]-glucosamine 3-sulfotransferase 1. |
| 2.7.9.2 Pyruvate, water dikinase. | 2.8.2.24 Desulfoglucosinolate sulfotransferase. |
| 2.7.9.3 Selenide, water dikinase. | 2.8.2.25 Flavonol 3-sulfotransferase. |
| 2.7.9.4 Alpha-glucan, water dikinase. | 2.8.2.26 Quercetin-3-sulfate 3'-sulfotransferase. |
| 2.8.1.1 Thiosulfate sulfur-transferase. | 2.8.2.27 Quercetin-3-sulfate 4'-sulfotransferase. |
| 2.8.1.2 3-mercaptopyruvate sulfur-transferase. | 2.8.2.28 Quercetin-3,3'-bissulfate 7-sulfotransferase. |
| 2.8.1.3 Thiosulfate--thiol sulfur-transferase. | 2.8.2.29 [Heparan sulfate]-glucosamine 3-sulfotransferase 2. |
| 2.8.1.4 tRNA sulfur-transferase. | 2.8.2.30 [Heparan sulfate]-glucosamine 3-sulfotransferase 3. |
| 2.8.1.5 Thiosulfate--dithiol sulfur-transferase. | 2.8.3.1 Propionate CoA-transferase. |
| 2.8.1.6 Biotin synthase. | 2.8.3.2 Oxalate CoA-transferase. |
| 2.8.1.7 Cysteine desulfurase. | 2.8.3.3 Malonate CoA-transferase. |

| | |
|---|---|
| 2.8.3.5 3-oxoacid CoA-transferase. | 3.1.1.34 Lipoprotein lipase. |
| 2.8.3.6 3-oxoadipate CoA-transferase. | 3.1.1.35 Dihydrocoumarin hydrolase. |
| 2.8.3.7 Succinate--citramalate CoA-transferase. | 3.1.1.36 Limonin-D-ring-lactonase. |
| 2.8.3.8 Acetate CoA-transferase. | 3.1.1.37 Steroid-lactonase. |
| 2.8.3.9 Butyrate--acetoacetate CoA-transferase. | 3.1.1.38 Triacetate-lactonase. |
| 2.8.3.10 Citrate CoA-transferase. | 3.1.1.39 Actinomycin lactonase. |
| 2.8.3.11 Citramalate CoA-transferase. | 3.1.1.40 Orsellinate-depside hydrolase. |
| 2.8.3.12 Glutaconate CoA-transferase. | 3.1.1.41 Cephalosporin-C deacetylase. |
| 2.8.3.13 Succinate--hydroxymethylglutarate CoA-transferase. | 3.1.1.42 Chlorogenate hydrolase. |
| 2.8.3.14 5-hydroxypentanoate CoA-transferase. | 3.1.1.43 Alpha-amino-acid esterase. |
| 2.8.3.15 Succinyl-CoA:(R)-benzylsuccinate CoA-transferase. | 3.1.1.44 4-methyloxaloacetate esterase. |
| 2.8.3.16 Formyl-CoA transferase. | 3.1.1.45 Carboxymethylenebutenolidase. |
| 2.8.3.17 Cinnamoyl-CoA:phenyllactate CoA-transferase. | 3.1.1.46 Deoxylimonate A-ring-lactonase. |
| 2.8.4.1 Coenzyme-B sulfoethylthiotransferase. | 3.1.1.47 1-alkyl-2-acetylglycerophosphocholine esterase. |
| 2.9.1.1 L-seryl-tRNA(Sec) selenium transferase. | 3.1.1.48 Fusarinine-C ornithinesterase. |
| | 3.1.1.49 Sinapine esterase. |
| ENZYME: 3.-.-.- | 3.1.1.50 Wax-ester hydrolase. |
| 3.1.1.1 Carboxylesterase. | 3.1.1.51 Phorbol-diester hydrolase. |
| 3.1.1.2 Arylesterase. | 3.1.1.52 Phosphatidylinositol deacylase. |
| 3.1.1.3 Triacylglycerol lipase. | 3.1.1.53 Sialate O-acetylesterase. |
| 3.1.1.4 Phospholipase A(2). | 3.1.1.54 Acetoxybutynylbithiophene deacetylase. |
| 3.1.1.5 Lysophospholipase. | 3.1.1.55 Acetylsalicylate deacetylase. |
| 3.1.1.6 Acetylesterase. | 3.1.1.56 Methylumbelliferyl-acetate deacetylase. |
| 3.1.1.7 Acetylcholinesterase. | 3.1.1.57 2-pyrone-4,6-dicarboxylate lactonase. |
| 3.1.1.8 Cholinesterase. | 3.1.1.58 N-acetylgalactosaminoglycan deacetylase. |
| 3.1.1.10 Tropinesterase. | 3.1.1.59 Juvenile-hormone esterase. |
| 3.1.1.11 Pectinesterase. | 3.1.1.60 Bis(2-ethylhexyl)phthalate esterase. |
| 3.1.1.13 Sterol esterase. | 3.1.1.61 Protein-glutamate methylesterase. |
| 3.1.1.14 Chlorophyllase. | 3.1.1.63 11-cis-retinyl-palmitate hydrolase. |

| | |
|---|---|
| 3.1.1.15 L-arabinonolactonase. | 3.1.1.64 All-trans-retinyl-palmitate hydrolase. |
| 3.1.1.17 Gluconolactonase. | 3.1.1.65 L-rhamnono-1,4-lactonase. |
| 3.1.1.19 Uronolactonase. | 3.1.1.66 5-(3,4-diacetoxybut-1-ynyl)-2,2'-bithiophene deacetylase. |
| 3.1.1.20 Tannase. | 3.1.1.67 Fatty-acyl-ethyl-ester synthase. |
| 3.1.1.21 Retinyl-palmitate esterase. | 3.1.1.68 Xylono-1,4-lactonase. |
| 3.1.1.22 Hydroxybutyrate-dimer hydrolase. | 3.1.1.70 Cetraxate benzylesterase. |
| 3.1.1.23 Acylglycerol lipase. | 3.1.1.71 Acetylalkylglycerol acetylhydrolase. |
| 3.1.1.24 3-oxoadipate enol-lactonase. | 3.1.1.72 Acetylxylan esterase. |
| 3.1.1.25 1,4-lactonase. | 3.1.1.73 Feruloyl esterase. |
| 3.1.1.26 Galactolipase. | 3.1.1.74 Cutinase. |
| 3.1.1.27 4-pyridoxolactonase. | 3.1.1.75 Poly(3-hydroxybutyrate) depolymerase. |
| 3.1.1.28 Acylcarnitine hydrolase. | 3.1.1.76 Poly(3-hydroxyoctanoate) depolymerase. |
| 3.1.1.29 Aminoacyl-tRNA hydrolase. | 3.1.1.77 Acyloxyacyl hydrolase. |
| 3.1.1.30 D-arabinonolactonase. | 3.1.1.78 Polyneuridine-aldehyde esterase. |
| 3.1.1.31 6-phosphogluconolactonase. | 3.1.1.79 Hormone-sensitive lipase. |
| 3.1.1.32 Phospholipase A(1). | 3.1.2.1 Acetyl-CoA hydrolase. |
| 3.1.1.33 6-acetylglucose deacetylase. | 3.1.2.2 Palmitoyl-CoA hydrolase. |

| | |
|---|---|
| 3.1.2.3 Succinyl-CoA hydrolase. | 3.1.3.23 Sugar-phosphatase. |
| 3.1.2.4 3-hydroxyisobutyryl-CoA hydrolase. | 3.1.3.24 Sucrose-phosphatase. |
| 3.1.2.5 Hydroxymethylglutaryl-CoA hydrolase. | 3.1.3.25 Inositol-1(or 4)-monophosphatase. |
| 3.1.2.6 Hydroxyacylglutathione hydrolase. | 3.1.3.26 4-phytase. |
| 3.1.2.7 Glutathione thiolesterase. | 3.1.3.27 Phosphatidylglycerophosphatase. |
| 3.1.2.10 Formyl-CoA hydrolase. | 3.1.3.28 ADP-phosphoglycerate phosphatase. |
| 3.1.2.11 Acetoacetyl-CoA hydrolase. | 3.1.3.29 N-acylneuraminate-9-phosphatase. |
| 3.1.2.12 S-formylglutathione hydrolase. | 3.1.3.31 Nucleotidase. |
| 3.1.2.13 S-succinylglutathione hydrolase. | 3.1.3.32 Polynucleotide 3'-phosphatase. |
| 3.1.2.14 Oleoyl-[acyl-carrier-protein] hydrolase. | 3.1.3.33 Polynucleotide 5'-phosphatase. |
| 3.1.2.15 Ubiquitin thiolesterase. | 3.1.3.34 Deoxynucleotide 3'-phosphatase. |
| 3.1.2.16 [Citrate-(pro-3S)-lyase] thiolesterase. | 3.1.3.35 Thymidylate 5'-phosphatase. |
| 3.1.2.17 (S)-methylmalonyl-CoA hydrolase. | 3.1.3.36 Phosphoinositide 5-phosphatase. |
| 3.1.2.18 ADP-dependent short-chain-acyl-CoA hydrolase. | 3.1.3.37 Sedoheptulose-bisphosphatase. |
| 3.1.2.19 ADP-dependent medium-chain-acyl-CoA hydrolase. | 3.1.3.38 3-phosphoglycerate phosphatase. |
| 3.1.2.20 Acyl-CoA hydrolase. | 3.1.3.39 Streptomycin-6-phosphatase. |
| 3.1.2.21 Dodecanoyl-[acyl-carrier protein] hydrolase. | 3.1.3.40 Guanidinodeoxy-scyllo-inositol-4-phosphatase. |
| 3.1.2.22 Palmitoyl-protein hydrolase. | 3.1.3.41 4-nitrophenylphosphatase. |
| 3.1.2.23 4-hydroxybenzoyl-CoA thioesterase. | 3.1.3.42 [Glycogen-synthase-D] phosphatase. |
| 3.1.2.24 2-(2-hydroxyphenyl)benzenesulfinate hydrolase. | 3.1.3.43 [Pyruvate dehydrogenase (lipoamide)]-phosphatase. |
| 3.1.2.25 Phenylacetyl-CoA hydrolase. | 3.1.3.44 [Acetyl-CoA carboxylase]-phosphatase. |
| 3.1.3.1 Alkaline phosphatase. | 3.1.3.45 3-deoxy-manno-octulosonate-8-phosphatase. |
| 3.1.3.2 Acid phosphatase. | 3.1.3.46 Fructose-2,6-bisphosphate 2-phosphatase. |
| 3.1.3.3 Phosphoserine phosphatase. | 3.1.3.47 [Hydroxymethylglutaryl-CoA reductase (NADPH)]-phosphatase. |
| 3.1.3.4 Phosphatidate phosphatase. | 3.1.3.48 Protein-tyrosine-phosphatase. |
| 3.1.3.5 5'-nucleotidase. | 3.1.3.49 [Pyruvate kinase]-phosphatase. |

| | |
|---|---|
| 3.1.3.6 3'-nucleotidase. | 3.1.3.50 Sorbitol-6-phosphatase. |
| 3.1.3.7 3'(2'),5'-bisphosphate nucleotidase. | 3.1.3.51 Dolichyl-phosphatase. |
| 3.1.3.8 3-phytase. | 3.1.3.52 [3-methyl-2-oxobutanoate dehydrogenase (lipoamide)]-phosphatase. |
| 3.1.3.9 Glucose-6-phosphatase. | 3.1.3.53 [Myosin light-chain]-phosphatase. |
| 3.1.3.10 Glucose-1-phosphatase. | 3.1.3.54 Fructose-2,6-bisphosphate 6-phosphatase. |
| 3.1.3.11 Fructose-bisphosphatase. | 3.1.3.55 Caldesmon-phosphatase. |
| 3.1.3.12 Trehalose-phosphatase. | 3.1.3.56 Inositol-polyphosphate 5-phosphatase. |
| 3.1.3.13 Bisphosphoglycerate phosphatase. | 3.1.3.57 Inositol-1,4-bisphosphate 1-phosphatase. |
| 3.1.3.14 Methylphosphothioglycerate phosphatase. | 3.1.3.58 Sugar-terminal-phosphatase. |
| 3.1.3.15 Histidinol-phosphatase. | 3.1.3.59 Alkylacetylglycerophosphatase. |
| 3.1.3.16 Phosphoprotein phosphatase. | 3.1.3.60 Phosphoenolpyruvate phosphatase. |
| 3.1.3.17 [Phosphorylase] phosphatase. | 3.1.3.62 Multiple inositol-polyphosphate phosphatase. |
| 3.1.3.18 Phosphoglycolate phosphatase. | 3.1.3.63 2-carboxy-D-arabinitol-1-phosphatase. |
| 3.1.3.19 Glycerol-2-phosphatase. | 3.1.3.64 Phosphatidylinositol-3-phosphatase. |
| 3.1.3.20 Phosphoglycerate phosphatase. | 3.1.3.66 Phosphatidylinositol-3,4-bisphosphate 4-phosphatase. |
| 3.1.3.21 Glycerol-1-phosphatase. | 3.1.3.67 Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase. |
| 3.1.3.22 Mannitol-1-phosphatase. | 3.1.3.68 2-deoxyglucose-6-phosphatase. |

| | |
|---|---|
| 3.1.3.69 Glucosylglycerol 3-phosphatase. | 3.1.13.3 Oligonucleotidase. |
| 3.1.3.70 Mannosyl-3-phosphoglycerate phosphatase. | 3.1.13.4 Poly(A)-specific ribonuclease. |
| 3.1.3.71 2-phosphosulfolactate phosphatase. | 3.1.14.1 Yeast ribonuclease. |
| 3.1.3.72 5-phytase. | 3.1.15.1 Venom exonuclease. |
| 3.1.3.73 Alpha-ribazole phosphatase. | 3.1.16.1 Spleen exonuclease. |
| 3.1.4.1 Phosphodiesterase I. | 3.1.21.1 Deoxyribonuclease I. |
| 3.1.4.2 Glycerophosphocholine phosphodiesterase. | 3.1.6.3 Glycosulfatase. |
| 3.1.4.3 Phospholipase C. | 3.1.6.4 N-acetylgalactosamine-6-sulfatase. |
| 3.1.4.4 Phospholipase D. | 3.1.6.6 Choline-sulfatase. |
| 3.1.4.11 Phosphoinositide phospholipase C. | 3.1.6.7 Cellulose-polysulfatase. |
| 3.1.4.12 Sphingomyelin phosphodiesterase. | 3.1.6.8 Cerebroside-sulfatase. |
| 3.1.4.13 Serine-ethanolaminephosphate phosphodiesterase. | 3.1.6.9 Chondro-4-sulfatase. |
| 3.1.4.14 [Acyl-carrier-protein] phosphodiesterase. | 3.1.6.10 Chondro-6-sulfatase. |
| 3.1.4.15 Adenylyl-[glutamate--ammonia ligase] hydrolase. | 3.1.6.11 Disulfoglucosamine-6-sulfatase. |
| 3.1.4.16 2',3'-cyclic-nucleotide 2'-phosphodiesterase. | 3.1.6.12 N-acetylgalactosamine-4-sulfatase. |
| 3.1.4.17 3',5'-cyclic-nucleotide phosphodiesterase. | 3.1.6.13 Iduronate-2-sulfatase. |
| 3.1.4.35 3',5'-cyclic-GMP phosphodiesterase. | 3.1.6.14 N-acetylglucosamine-6-sulfatase. |
| 3.1.4.37 2',3'-cyclic-nucleotide 3'-phosphodiesterase. | 3.1.6.15 N-sulfoglucosamine-3-sulfatase. |
| 3.1.4.38 Glycerophosphocholine cholinephosphodiesterase. | 3.1.6.16 Monomethyl-sulfatase. |
| 3.1.4.39 Alkylglycerophosphoethanolamine phosphodiesterase. | 3.1.6.17 D-lactate-2-sulfatase. |

| | |
|---|---|
| 3.1.4.40 CMP-N-acylneuraminate phosphodiesterase. | 3.1.6.18 Glucuronate-2-sulfatase. |
| 3.1.4.41 Sphingomyelin phosphodiesterase D. | 3.1.7.1 Prenyl-diphosphatase. |
| 3.1.4.42 Glycerol-1,2-cyclic-phosphate 2-phosphodiesterase. | 3.1.7.2 Guanosine-3',5'-bis(diphosphate) 3'-diphosphatase. |
| 3.1.4.43 Glycerophosphoinositol inositolphosphodiesterase. | 3.1.7.3 Monoterpenyl-diphosphatase. |
| 3.1.4.44 Glycerophosphoinositol glycerophosphodiesterase. | 3.1.8.1 Aryldialkylphosphatase. |
| 3.1.4.45 N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. | 3.1.8.2 Diisopropyl-fluorophosphatase. |
| 3.1.4.46 Glycerophosphodiester phosphodiesterase. | 3.1.11.1 Exodeoxyribonuclease I. |
| 3.1.4.48 Dolichylphosphate-glucose phosphodiesterase. | 3.1.11.2 Exodeoxyribonuclease III. |
| 3.1.4.49 Dolichylphosphate-mannose phosphodiesterase. | 3.1.11.3 Exodeoxyribonuclease (lambda-induced). |
| 3.1.4.50 Glycosylphosphatidylinositol phospholipase D. | 3.1.11.4 Exodeoxyribonuclease (phage Sp3-induced). |
| 3.1.4.51 Glucose-1-phospho-D-mannosylglycoprotein phosphodiesterase. | 3.1.11.5 Exodeoxyribonuclease V. |
| 3.1.5.1 dGTPase. | 3.1.11.6 Exodeoxyribonuclease VII. |
| 3.1.6.1 Arylsulfatase. | 3.1.13.1 Exoribonuclease II. |
| 3.1.6.2 Steryl-sulfatase. | 3.1.13.2 Exoribonuclease H. |

| | |
|---|---|
| 3.1.21.2 Deoxyribonuclease IV (phage-T(4)-induced). | 3.2.1.23 Beta-galactosidase. |
| 3.1.21.3 Type I site-specific deoxyribonuclease. | 3.2.1.24 Alpha-mannosidase. |
| 3.1.21.4 Type II site-specific deoxyribonuclease. | 3.2.1.25 Beta-mannosidase. |
| 3.1.21.5 Type III site-specific deoxyribonuclease. | 3.2.1.26 Beta-fructofuranosidase. |
| 3.1.21.6 CC-preferring endodeoxyribonuclease. | 3.2.1.28 Alpha,alpha-trehalase. |
| 3.1.21.7 Deoxyribonuclease V. | 3.2.1.31 Beta-glucuronidase. |
| 3.1.22.1 Deoxyribonuclease II. | 3.2.1.32 Xylan endo-1,3-beta-xylosidase. |
| 3.1.22.2 Aspergillus deoxyribonuclease K(1). | 3.2.1.33 Amylo-alpha-1,6-glucosidase. |
| 3.1.22.4 Crossover junction endoribonuclease. | 3.2.1.35 Hyaluronoglucosaminidase. |
| 3.1.22.5 Deoxyribonuclease X. | 3.2.1.36 Hyaluronoglucuronidase. |
| 3.1.25.1 Deoxyribonuclease (pyrimidine dimer). | 3.2.1.37 Xylan 1,4-beta-xylosidase. |
| 3.1.26.1 Physarum polycephalum ribonuclease. | 3.2.1.38 Beta-D-fucosidase. |
| 3.1.26.2 Ribonuclease alpha. | 3.2.1.39 Glucan endo-1,3-beta-D-glucosidase. |
| 3.1.26.3 Ribonuclease III. | 3.2.1.40 Alpha-L-rhamnosidase. |
| 3.1.26.4 Ribonuclease H. | 3.2.1.41 Pullulanase. |
| 3.1.26.5 Ribonuclease P. | 3.2.1.42 GDP-glucosidase. |
| 3.1.26.6 Ribonuclease IV. | 3.2.1.43 Beta-L-rhamnosidase. |
| 3.1.26.7 Ribonuclease P4. | 3.2.1.44 Fucoidanase. |
| 3.1.26.8 Ribonuclease M5. | 3.2.1.45 Glucosylceramidase. |
| 3.1.26.9 Ribonuclease (poly-(U)-specific). | 3.2.1.46 Galactosylceramidase. |
| 3.1.26.10 Ribonuclease IX. | 3.2.1.47 Galactosylgalactosylglucosylceramidase. |
| 3.1.26.11 Ribonuclease Z. | 3.2.1.48 Sucrose alpha-glucosidase. |
| 3.1.27.1 Ribonuclease T(2). | 3.2.1.49 Alpha-N-acetylgalactosaminidase. |
| 3.1.27.2 Bacillus subtilis ribonuclease. | 3.2.1.50 Alpha-N-acetylglucosaminidase. |
| 3.1.27.3 Ribonuclease T(1). | 3.2.1.51 Alpha-L-fucosidase. |
| 3.1.27.4 Ribonuclease U(2). | 3.2.1.52 Beta-N-acetylhexosaminidase. |
| 3.1.27.5 Pancreatic ribonuclease. | 3.2.1.53 Beta-N-acetylgalactosaminidase. |

| | |
|---|---|
| 3.1.27.6 Enterobacter ribonuclease. | 3.2.1.54 Cyclomaltodextrinase. |
| 3.1.27.7 Ribonuclease F. | 3.2.1.55 Alpha-N-arabinofuranosidase. |
| 3.1.27.8 Ribonuclease V. | 3.2.1.56 Glucuronosyl-disulfoglucosamine glucuronidase. |
| 3.1.27.9 tRNA-intron endonuclease. | 3.2.1.57 Isopullulanase. |
| 3.1.27.10 rRNA endonuclease. | 3.2.1.58 Glucan 1,3-beta-glucosidase. |
| 3.1.30.1 Aspergillus nuclease S(1). | 3.2.1.59 Glucan endo-1,3-alpha-glucosidase. |
| 3.1.30.2 Serratia marcescens nuclease. | 3.2.1.60 Glucan 1,4-alpha-maltotetraohydrolase. |
| 3.1.31.1 Micrococcal nuclease. | 3.2.1.61 Mycodextranase. |
| 3.2.1.1 Alpha-amylase. | 3.2.1.62 Glycosylceramidase. |
| 3.2.1.2 Beta-amylase. | 3.2.1.63 1,2-alpha-L-fucosidase. |
| 3.2.1.3 Glucan 1,4-alpha-glucosidase. | 3.2.1.64 2,6-beta-fructan 6-levanbiohydrolase. |
| 3.2.1.4 Cellulase. | 3.2.1.65 Levanase. |
| 3.2.1.6 Endo-1,3(4)-beta-glucanase. | 3.2.1.66 Quercitrinase. |
| 3.2.1.7 Inulinase. | 3.2.1.67 Galacturan 1,4-alpha-galacturonidase. |
| 3.2.1.8 Endo-1,4-beta-xylanase. | 3.2.1.68 Isoamylase. |
| 3.2.1.10 Oligo-1,6-glucosidase. | 3.2.1.70 Glucan 1,6-alpha-glucosidase. |
| 3.2.1.11 Dextranase. | 3.2.1.71 Glucan endo-1,2-beta-glucosidase. |
| 3.2.1.14 Chitinase. | 3.2.1.72 Xylan 1,3-beta-xylosidase. |
| 3.2.1.15 Polygalacturonase. | 3.2.1.73 Licheninase. |
| 3.2.1.17 Lysozyme. | 3.2.1.74 Glucan 1,4-beta-glucosidase. |
| 3.2.1.18 Exo-alpha-sialidase. | 3.2.1.75 Glucan endo-1,6-beta-glucosidase. |
| 3.2.1.20 Alpha-glucosidase. | 3.2.1.76 L-iduronidase. |
| 3.2.1.21 Beta-glucosidase. | 3.2.1.77 Mannan 1,2-(1,3)-alpha-mannosidase. |
| 3.2.1.22 Alpha-galactosidase. | 3.2.1.78 Mannan endo-1,4-beta-mannosidase. |

| | |
|---|---|
| 3.2.1.80 Fructan beta-fructosidase. | 3.2.1.121 Polymannuronate hydrolase. |
| 3.2.1.81 Agarase. | 3.2.1.122 Maltose-6'-phosphate glucosidase. |
| 3.2.1.82 Exo-poly-alpha-galacturonosidase. | 3.2.1.123 Endoglycosylceramidase. |
| 3.2.1.83 Kappa-carrageenase. | 3.2.1.124 3-deoxy-2-octulosonidase. |
| 3.2.1.84 Glucan 1,3-alpha-glucosidase. | 3.2.1.125 Raucaffricine beta-glucosidase. |
| 3.2.1.85 6-phospho-beta-galactosidase. | 3.2.1.126 Coniferin beta-glucosidase. |
| 3.2.1.86 6-phospho-beta-glucosidase. | 3.2.1.127 1,6-alpha-L-fucosidase. |
| 3.2.1.87 Capsular-polysaccharide endo-1,3-alpha-galactosidase. | 3.2.1.128 Glycyrrhizinate beta-glucuronidase. |
| 3.2.1.88 Beta-L-arabinosidase. | 3.2.1.129 Endo-alpha-sialidase. |
| 3.2.1.89 Arabinogalactan endo-1,4-beta-galactosidase. | 3.2.1.130 Glycoprotein endo-alpha-1,2-mannosidase. |
| 3.2.1.91 Cellulose 1,4-beta-cellobiosidase. | 3.2.1.131 Xylan alpha-1,2-glucuronosidase. |
| 3.2.1.92 Peptidoglycan beta-N-acetylmuramidase. | 3.2.1.132 Chitosanase. |
| 3.2.1.93 Alpha,alpha-phosphotrehalase. | 3.2.1.133 Glucan 1,4-alpha-maltohydrolase. |
| 3.2.1.94 Glucan 1,6-alpha-isomaltosidase. | 3.2.1.134 Difructose-anhydride synthase. |
| 3.2.1.95 Dextran 1,6-alpha-isomaltotriosidase. | 3.2.1.135 Neopullulanase. |
| 3.2.1.96 Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase. | 3.2.1.136 Glucuronoarabinoxylan endo-1,4-beta-xylanase. |
| 3.2.1.97 Glycopeptide alpha-N-acetylgalactosaminidase. | 3.2.1.137 Mannan exo-1,2-1,6-alpha-mannosidase. |
| 3.2.1.98 Glucan 1,4-alpha-maltohexaosidase. | 3.2.1.139 Alpha-glucuronidase. |
| 3.2.1.99 Arabinan endo-1,5-alpha-L-arabinosidase. | 3.2.1.140 Lacto-N-biosidase. |
| 3.2.1.100 Mannan 1,4-mannobiosidase. | 3.2.1.141 4-alpha-D-{(1->4)-alpha-D-glucano}trehalose trehalohydrolase. |
| 3.2.1.101 Mannan endo-1,6-alpha-mannosidase. | 3.2.1.142 Limit dextrinase. |

| | |
|---|---|
| 3.2.1.102 Blood-group-substance endo-1,4-beta-galactosidase. | 3.2.1.143 Poly(ADP-ribose) glycohydrolase. |
| 3.2.1.103 Keratan-sulfate endo-1,4-beta-galactosidase. | 3.2.1.144 3-deoxyoctulosonase. |
| 3.2.1.104 Steryl-beta-glucosidase. | 3.2.1.145 Galactan 1,3-beta-galactosidase. |
| 3.2.1.105 Strictosidine beta-glucosidase. | 3.2.1.146 Beta-galactofuranosidase. |
| 3.2.1.106 Mannosyl-oligosaccharide glucosidase. | 3.2.1.147 Thioglucosidase. |
| 3.2.1.107 Protein-glucosylgalactosylhydroxylysine glucosidase. | 3.2.1.148 Ribosylhomocysteinase. |
| 3.2.1.108 Lactase. | 3.2.1.149 Beta-primeverosidase. |
| 3.2.1.109 Endogalactosaminidase. | 3.2.1.150 Oligoxyloglucan reducing-end-specific cellobiohydrolase. |
| 3.2.1.110 Mucinaminylserine mucinaminidase. | 3.2.1.151 Xyloglucan-specific endo-beta-1,4-glucanase. |
| 3.2.1.111 1,3-alpha-L-fucosidase. | 3.2.2.1 Purine nucleosidase. |
| 3.2.1.112 2-deoxyglucosidase. | 3.2.2.2 Inosine nucleosidase. |
| 3.2.1.113 Mannosyl-oligosaccharide 1,2-alpha-mannosidase. | 3.2.2.3 Uridine nucleosidase. |
| 3.2.1.114 Mannosyl-oligosaccharide 1,3-1,6-alpha-mannosidase. | 3.2.2.4 AMP nucleosidase. |
| 3.2.1.115 Branched-dextran exo-1,2-alpha-glucosidase. | 3.2.2.5 NAD(+) nucleosidase. |
| 3.2.1.116 Glucan 1,4-alpha-maltotriohydrolase. | 3.2.2.6 NAD(P)(+) nucleosidase. |
| 3.2.1.117 Amygdalin beta-glucosidase. | 3.2.2.7 Adenosine nucleosidase. |
| 3.2.1.118 Prunasin beta-glucosidase. | 3.2.2.8 Ribosylpyrimidine nucleosidase. |
| 3.2.1.119 Vicianin beta-glucosidase. | 3.2.2.9 Adenosylhomocysteine nucleosidase. |
| 3.2.1.120 Oligoxyloglucan beta-glycosidase. | 3.2.2.10 Pyrimidine-5'-nucleotide nucleosidase. |

| | |
|---|---|
| 3.2.2.11 Beta-aspartyl-N-acetylglucosaminidase. | 3.4.13.18 Cytosol nonspecific dipeptidase. |
| 3.2.2.12 Inosinate nucleosidase. | 3.4.13.19 Membrane dipeptidase. |
| 3.2.2.13 1-methyladenosine nucleosidase. | 3.4.13.20 Beta-Ala-His dipeptidase. |
| 3.2.2.14 NMN nucleosidase. | 3.4.13.21 Dipeptidase E. |
| 3.2.2.15 DNA-deoxyinosine glycosylase. | 3.4.14.1 Dipeptidyl-peptidase I. |
| 3.2.2.16 Methylthioadenosine nucleosidase. | 3.4.14.2 Dipeptidyl-peptidase II. |
| 3.2.2.17 Deoxyribodipyrimidine endonucleosidase. | 3.4.14.4 Dipeptidyl-peptidase III. |
| 3.2.2.19 [Protein ADP-ribosylarginine] hydrolase. | 3.4.14.5 Dipeptidyl-peptidase IV. |
| 3.2.2.20 DNA-3-methyladenine glycosylase I. | 3.4.14.6 Dipeptidyl-dipeptidase. |
| 3.2.2.21 DNA-3-methyladenine glycosylase II. | 3.4.14.9 Tripeptidyl-peptidase I. |
| 3.2.2.22 rRNA N-glycosylase. | 3.4.14.10 Tripeptidyl-peptidase II. |
| 3.2.2.23 DNA-formamidopyrimidine glycosylase. | 3.4.14.11 Xaa-Pro dipeptidyl-peptidase. |
| 3.2.2.24 ADP-ribosyl-[dinitrogen reductase] hydrolase. | 3.4.15.1 Peptidyl-dipeptidase A. |
| 3.3.1.1 Adenosylhomocysteinase. | 3.4.15.4 Peptidyl-dipeptidase B. |
| 3.3.1.2 Adenosylmethionine hydrolase. | 3.4.15.5 Peptidyl-dipeptidase Dcp. |
| 3.3.2.1 Isochorismatase. | 3.4.16.2 Lysosomal Pro-X carboxypeptidase. |
| 3.3.2.2 Alkenylglycerophosphocholine hydrolase. | 3.4.16.4 Serine-type D-Ala-D-Ala carboxypeptidase. |
| 3.3.2.3 Epoxide hydrolase. | 3.4.16.5 Carboxypeptidase C. |
| 3.3.2.4 Trans-epoxysuccinate hydrolase. | 3.4.16.6 Carboxypeptidase D. |
| 3.3.2.5 Alkenylglycerophosphoethanolamine hydrolase. | 3.4.17.1 Carboxypeptidase A. |
| 3.3.2.6 Leukotriene-A(4) hydrolase. | 3.4.17.2 Carboxypeptidase B. |
| 3.3.2.7 Hepoxilin-epoxide hydrolase. | 3.4.17.3 Lysine carboxypeptidase. |

| | |
|---|---|
| 3.3.2.8 Limonene-1,2-epoxide hydrolase. | 3.4.17.4 Gly-X carboxypeptidase. |
| 3.4.11.1 Leucyl aminopeptidase. | 3.4.17.6 Alanine carboxypeptidase. |
| 3.4.11.2 Membrane alanyl aminopeptidase. | 3.4.17.8 Muramoylpentapeptide carboxypeptidase. |
| 3.4.11.3 Cystinyl aminopeptidase. | 3.4.17.10 Carboxypeptidase E. |
| 3.4.11.4 Tripeptide aminopeptidase. | 3.4.17.11 Glutamate carboxypeptidase. |
| 3.4.11.5 Prolyl aminopeptidase. | 3.4.17.12 Carboxypeptidase M. |
| 3.4.11.6 Aminopeptidase B. | 3.4.17.13 Muramoyltetrapeptide carboxypeptidase. |
| 3.4.11.7 Glutamyl aminopeptidase. | 3.4.17.14 Zinc D-Ala-D-Ala carboxypeptidase. |
| 3.4.11.9 Xaa-Pro aminopeptidase. | 3.4.17.15 Carboxypeptidase A2. |
| 3.4.11.10 Bacterial leucyl aminopeptidase. | 3.4.17.16 Membrane Pro-X carboxypeptidase. |
| 3.4.11.13 Clostridial aminopeptidase. | 3.4.17.17 Tubulinyl-Tyr carboxypeptidase. |
| 3.4.11.14 Cytosol alanyl aminopeptidase. | 3.4.17.18 Carboxypeptidase T. |
| 3.4.11.15 Aminopeptidase Y. | 3.4.17.19 Carboxypeptidase Taq. |
| 3.4.11.16 Xaa-Trp aminopeptidase. | 3.4.17.20 Carboxypeptidase U. |
| 3.4.11.17 Tryptophanyl aminopeptidase. | 3.4.17.21 Glutamate carboxypeptidase II. |
| 3.4.11.18 Methionyl aminopeptidase. | 3.4.17.22 Metallocarboxypeptidase D. |
| 3.4.11.19 D-stereospecific aminopeptidase. | 3.4.18.1 Cathepsin X. |
| 3.4.11.20 Aminopeptidase Ey. | 3.4.19.1 Acylaminoacyl-peptidase. |
| 3.4.11.21 Aspartyl aminopeptidase. | 3.4.19.2 Peptidyl-glycinamidase. |
| 3.4.11.22 Aminopeptidase I. | 3.4.19.3 Pyroglutamyl-peptidase I. |
| 3.4.11.23 PepB aminopeptidase. | 3.4.19.5 Beta-aspartyl-peptidase. |
| 3.4.13.3 Xaa-His dipeptidase. | 3.4.19.6 Pyroglutamyl-peptidase II. |
| 3.4.13.4 Xaa-Arg dipeptidase. | 3.4.19.7 N-formylmethionyl-peptidase. |
| 3.4.13.5 Xaa-methyl-His dipeptidase. | 3.4.19.9 Gamma-glutamyl hydrolase. |
| 3.4.13.7 Glu-Glu dipeptidase. | 3.4.19.11 Gamma-D-glutamyl-meso-diaminopimelate peptidase. |
| 3.4.13.9 Xaa-Pro dipeptidase. | 3.4.19.12 Ubiquitinyl hydrolase 1. |
| 3.4.13.12 Met-Xaa dipeptidase. | 3.4.21.1 Chymotrypsin. |
| 3.4.13.17 Non-stereospecific dipeptidase. | 3.4.21.2 Chymotrypsin C. |

| | |
|---|---|
| 3.4.21.3 Metridin. | 3.4.21.74 Venombin A. |
| 3.4.21.4 Trypsin. | 3.4.21.75 Furin. |
| 3.4.21.5 Thrombin. | 3.4.21.76 Myeloblastin. |
| 3.4.21.6 Coagulation factor Xa. | 3.4.21.77 Semenogelase. |
| 3.4.21.7 Plasmin. | 3.4.21.78 Granzyme A. |
| 3.4.21.9 Enteropeptidase. | 3.4.21.79 Granzyme B. |
| 3.4.21.10 Acrosin. | 3.4.21.80 Streptogrisin A. |
| 3.4.21.12 Alpha-lytic endopeptidase. | 3.4.21.81 Streptogrisin B. |
| 3.4.21.19 Glutamyl endopeptidase. | 3.4.21.82 Glutamyl endopeptidase II. |
| 3.4.21.20 Cathepsin G. | 3.4.21.83 Oligopeptidase B. |
| 3.4.21.21 Coagulation factor VIIa. | 3.4.21.84 Limulus clotting factor C. |
| 3.4.21.22 Coagulation factor IXa. | 3.4.21.85 Limulus clotting factor B. |
| 3.4.21.25 Cucumisin. | 3.4.21.86 Limulus clotting enzyme. |
| 3.4.21.26 Prolyl oligopeptidase. | 3.4.21.87 Omptin. |
| 3.4.21.27 Coagulation factor XIa. | 3.4.21.88 Repressor lexA. |
| 3.4.21.32 Brachyurin. | 3.4.21.89 Signal peptidase I. |
| 3.4.21.34 Plasma kallikrein. | 3.4.21.90 Togavirin. |
| 3.4.21.35 Tissue kallikrein. | 3.4.21.91 Flavivirin. |
| 3.4.21.36 Pancreatic elastase. | 3.4.21.92 Endopeptidase Clp. |
| 3.4.21.37 Leukocyte elastase. | 3.4.21.93 Proprotein convertase 1. |
| 3.4.21.38 Coagulation factor XIIa. | 3.4.21.94 Proprotein convertase 2. |

| | |
|---|---|
| 3.4.21.39 Chymase. | 3.4.21.95 Snake venom factor V activator. |
| 3.4.21.41 Complement subcomponent C1r. | 3.4.21.96 Lactocepin. |
| 3.4.21.42 Complement subcomponent C1s. | 3.4.21.97 Assemblin. |
| 3.4.21.43 Classical-complement-pathway C3/C5 convertase. | 3.4.21.98 Hepacivirin. |
| 3.4.21.45 Complement factor I. | 3.4.21.99 Spermosin. |
| 3.4.21.46 Complement factor D. | 3.4.21.100 Pseudomonalisin. |
| 3.4.21.47 Alternative-complement-pathway C3/C5 convertase. | 3.4.21.101 Xanthomonalisin. |
| 3.4.21.48 Cerevisin. | 3.4.21.102 C-terminal processing peptidase. |
| 3.4.21.49 Hypodermin C. | 3.4.21.103 Physarolisin. |
| 3.4.21.50 Lysyl endopeptidase. | 3.4.22.1 Cathepsin B. |
| 3.4.21.53 Endopeptidase La. | 3.4.22.2 Papain. |
| 3.4.21.54 Gamma-renin. | 3.4.22.3 Ficain. |
| 3.4.21.55 Venombin AB. | 3.4.22.6 Chymopapain. |
| 3.4.21.57 Leucyl endopeptidase. | 3.4.22.7 Asclepain. |
| 3.4.21.59 Tryptase. | 3.4.22.8 Clostripain. |
| 3.4.21.60 Scutelarin. | 3.4.22.10 Streptopain. |
| 3.4.21.61 Kexin. | 3.4.22.14 Actinidain. |
| 3.4.21.62 Subtilisin. | 3.4.22.15 Cathepsin L. |
| 3.4.21.63 Oryzin. | 3.4.22.16 Cathepsin H. |
| 3.4.21.64 Endopeptidase K. | 3.4.22.24 Cathepsin T. |
| 3.4.21.65 Thermomycolin. | 3.4.22.25 Glycyl endopeptidase. |
| 3.4.21.66 Thermitase. | 3.4.22.26 Cancer procoagulant. |
| 3.4.21.67 Endopeptidase So. | 3.4.22.27 Cathepsin S. |
| 3.4.21.68 T-plasminogen activator. | 3.4.22.28 Picornain 3C. |
| 3.4.21.69 Protein C (activated). | 3.4.22.29 Picornain 2A. |
| 3.4.21.70 Pancreatic endopeptidase E. | 3.4.22.30 Caricain. |
| 3.4.21.71 Pancreatic elastase II. | 3.4.22.31 Ananain. |
| 3.4.21.72 IgA-specific serine endopeptidase. | 3.4.22.32 Stem bromelain. |
| 3.4.21.73 U-plasminogen activator. | 3.4.22.33 Fruit bromelain. |

| | |
|---|---|
| 3.4.22.34 Legumain. | 3.4.23.44 Nodavirus endopeptidase. |
| 3.4.22.35 Histolysain. | 3.4.23.45 Memapsin 1. |
| 3.4.22.36 Caspase-1. | 3.4.23.46 Memapsin 2. |
| 3.4.22.37 Gingipain R. | 3.4.23.47 HIV-2 retropepsin. |
| 3.4.22.38 Cathepsin K. | 3.4.23.48 Plasminogen activator Pla. |
| 3.4.22.39 Adenain. | 3.4.24.1 Atrolysin A. |
| 3.4.22.40 Bleomycin hydrolase. | 3.4.24.3 Microbial collagenase. |
| 3.4.22.41 Cathepsin F. | 3.4.24.6 Leucolysin. |
| 3.4.22.42 Cathepsin O. | 3.4.24.7 Interstitial collagenase. |
| 3.4.22.43 Cathepsin V. | 3.4.24.11 Neprilysin. |
| 3.4.22.44 Nuclear-inclusion-a endopeptidase. | 3.4.24.12 Envelysin. |
| 3.4.22.45 Helper-component proteinase. | 3.4.24.13 IgA-specific metalloendopeptidase. |
| 3.4.22.46 L-peptidase. | 3.4.24.14 Procollagen N-endopeptidase. |
| 3.4.22.47 Gingipain K. | 3.4.24.15 Thimet oligopeptidase. |
| 3.4.22.48 Staphopain. | 3.4.24.16 Neurolysin. |
| 3.4.22.49 Separase. | 3.4.24.17 Stromelysin 1. |
| 3.4.22.50 V-cath endopeptidase. | 3.4.24.18 Meprin A. |
| 3.4.22.51 Cruzipain. | 3.4.24.19 Procollagen C-endopeptidase. |
| 3.4.22.52 Calpain-1. | 3.4.24.20 Peptidyl-Lys metalloendopeptidase. |
| 3.4.22.53 Calpain-2. | 3.4.24.21 Astacin. |
| 3.4.23.1 Pepsin A. | 3.4.24.22 Stromelysin 2. |

| | |
|---|---|
| 3.4.23.2 Pepsin B. | 3.4.24.23 Matrilysin. |
| 3.4.23.3 Gastricsin. | 3.4.24.24 Gelatinase A. |
| 3.4.23.4 Chymosin. | 3.4.24.25 Vibriolysin. |
| 3.4.23.5 Cathepsin D. | 3.4.24.26 Pseudolysin. |
| 3.4.23.12 Nepenthesin. | 3.4.24.27 Thermolysin. |
| 3.4.23.15 Renin. | 3.4.24.28 Bacillolysin. |
| 3.4.23.16 HIV-1 retropepsin. | 3.4.24.29 Aureolysin. |
| 3.4.23.17 Pro-opiomelanocortin converting enzyme. | 3.4.24.30 Coccolysin. |
| 3.4.23.18 Aspergillopepsin I. | 3.4.24.31 Mycolysin. |
| 3.4.23.19 Aspergillopepsin II. | 3.4.24.32 Beta-lytic metalloendopeptidase. |
| 3.4.23.20 Penicillopepsin. | 3.4.24.33 Peptidyl-Asp metalloendopeptidase. |
| 3.4.23.21 Rhizopuspepsin. | 3.4.24.34 Neutrophil collagenase. |
| 3.4.23.22 Endothiapepsin. | 3.4.24.35 Gelatinase B. |
| 3.4.23.23 Mucorpepsin. | 3.4.24.36 Leishmanolysin. |
| 3.4.23.24 Candidapepsin. | 3.4.24.37 Saccharolysin. |
| 3.4.23.25 Saccharopepsin. | 3.4.24.38 Gametolysin. |
| 3.4.23.26 Rhodotorulapepsin. | 3.4.24.39 Deuterolysin. |
| 3.4.23.28 Acrocylindropepsin. | 3.4.24.40 Serralysin. |
| 3.4.23.29 Polyporopepsin. | 3.4.24.41 Atrolysin B. |
| 3.4.23.30 Pycnoporopepsin. | 3.4.24.42 Atrolysin C. |
| 3.4.23.31 Scytalidopepsin A. | 3.4.24.43 Atroxase. |
| 3.4.23.32 Scytalidopepsin B. | 3.4.24.44 Atrolysin E. |
| 3.4.23.34 Cathepsin E. | 3.4.24.45 Atrolysin F. |
| 3.4.23.35 Barrierpepsin. | 3.4.24.46 Adamalysin. |
| 3.4.23.36 Signal peptidase II. | 3.4.24.47 Horrilysin. |
| 3.4.23.38 Plasmepsin I. | 3.4.24.48 Ruberlysin. |
| 3.4.23.39 Plasmepsin II. | 3.4.24.49 Bothropasin. |
| 3.4.23.40 Phytepsin. | 3.4.24.50 Bothrolysin. |
| 3.4.23.41 Yapsin 1. | 3.4.24.51 Ophiolysin. |
| 3.4.23.42 Thermopsin. | 3.4.24.52 Trimerelysin I. |
| 3.4.23.43 Prepilin peptidase. | 3.4.24.53 Trimerelysin II. |

| | |
|---|---|
| 3.4.24.54 Mucrolysin. | 3.5.1.16 Acetylornithine deacetylase. |
| 3.4.24.55 Pitrilysin. | 3.5.1.17 Acyl-lysine deacylase. |
| 3.4.24.56 Insulysin. | 3.5.1.18 Succinyl-diaminopimelate desuccinylase. |
| 3.4.24.57 O-sialoglycoprotein endopeptidase. | 3.5.1.19 Nicotinamidase. |
| 3.4.24.58 Russellysin. | 3.5.1.20 Citrullinase. |
| 3.4.24.59 Mitochondrial intermediate peptidase. | 3.5.1.21 N-acetyl-beta-alanine deacetylase. |
| 3.4.24.60 Dactylysin. | 3.5.1.22 Pantothenase. |
| 3.4.24.61 Nardilysin. | 3.5.1.23 Ceramidase. |
| 3.4.24.62 Magnolysin. | 3.5.1.24 Choloylglycine hydrolase. |
| 3.4.24.63 Meprin B. | 3.5.1.25 N-acetylglucosamine-6-phosphate deacetylase. |
| 3.4.24.64 Mitochondrial processing peptidase. | 3.5.1.26 N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase. |
| 3.4.24.65 Macrophage elastase. | 3.5.1.27 N-formylmethionylaminoacyl-tRNA deformylase. |
| 3.4.24.66 Choriolysin L. | 3.5.1.28 N-acetylmuramoyl-L-alanine amidase. |
| 3.4.24.67 Choriolysin H. | 3.5.1.29 2-(acetamidomethylene)succinate hydrolase. |
| 3.4.24.68 Tentoxilysin. | 3.5.1.30 5-aminopentanamidase. |
| 3.4.24.69 Bontoxilysin. | 3.5.1.31 Formylmethionine deformylase. |
| 3.4.24.70 Oligopeptidase A. | 3.5.1.32 Hippurate hydrolase. |

| | |
|---|---|
| 3.4.24.71 Endothelin-converting enzyme 1. | 3.5.1.33 N-acetylglucosamine deacetylase. |
| 3.4.24.72 Fibrolase. | 3.5.1.35 D-glutaminase. |
| 3.4.24.73 Jararhagin. | 3.5.1.36 N-methyl-2-oxoglutaramate hydrolase. |
| 3.4.24.74 Fragilysin. | 3.5.1.38 Glutamin-(asparagin-)ase. |
| 3.4.24.75 Lysostaphin. | 3.5.1.39 Alkylamidase. |
| 3.4.24.76 Flavastacin. | 3.5.1.40 Acylagmatine amidase. |
| 3.4.24.77 Snapalysin. | 3.5.1.41 Chitin deacetylase. |
| 3.4.24.78 GPR endopeptidase. | 3.5.1.42 Nicotinamide-nucleotide amidase. |
| 3.4.24.79 Pappalysin-1. | 3.5.1.43 Peptidyl-glutaminase. |
| 3.4.24.80 Membrane-type matrix metalloproteinase-1. | 3.5.1.44 Protein-glutamine glutaminase. |
| 3.4.24.81 ADAM10 endopeptidase. | 3.5.1.46 6-aminohexanoate-dimer hydrolase. |
| 3.4.24.82 ADAMTS-4 endopeptidase. | 3.5.1.47 N-acetyldiaminopimelate deacetylase. |
| 3.4.24.83 Anthrax lethal factor endopeptidase. | 3.5.1.48 Acetylspermidine deacetylase. |
| 3.4.24.84 Ste24 endopeptidase. | 3.5.1.49 Formamidase. |
| 3.4.24.85 S2P endopeptidase. | 3.5.1.50 Pentanamidase. |
| 3.4.24.86 ADAM 17 endopeptidase. | 3.5.1.51 4-acetamidobutyryl-CoA deacetylase. |
| 3.4.25.1 Proteasome endopeptidase complex. | 3.5.1.52 Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase. |
| 3.5.1.1 Asparaginase. | 3.5.1.53 N-carbamoylputrescine amidase. |
| 3.5.1.2 Glutaminase. | 3.5.1.54 Allophanate hydrolase. |
| 3.5.1.3 Omega-amidase. | 3.5.1.55 Long-chain-fatty-acyl-glutamate deacylase. |
| 3.5.1.4 Amidase. | 3.5.1.56 N,N-dimethylformamidase. |
| 3.5.1.5 Urease. | 3.5.1.57 Tryptophanamidase. |
| 3.5.1.6 Beta-ureidopropionase. | 3.5.1.58 N-benzyloxycarbonylglycine hydrolase. |
| 3.5.1.7 Ureidosuccinase. | 3.5.1.59 N-carbamoylsarcosine amidase. |
| 3.5.1.8 Formylaspartate deformylase. | 3.5.1.60 N-(long-chain-acyl)ethanolamine deacylase. |
| 3.5.1.9 Arylformamidase. | 3.5.1.61 Mimosinase. |
| 3.5.1.10 Formyltetrahydrofolate deformylase. | 3.5.1.62 Acetylputrescine deacetylase. |
| 3.5.1.11 Penicillin amidase. | 3.5.1.63 4-acetamidobutyrate deacetylase. |
| 3.5.1.12 Biotinidase. | 3.5.1.64 N(alpha)-benzyloxycarbonylleucine hydrolase. |
| 3.5.1.13 Aryl-acylamidase. | 3.5.1.65 Theanine hydrolase. |
| 3.5.1.14 Aminoacylase. | 3.5.1.66 2-(hydroxymethyl)-3-(acetamidomethylene)succinate hydrolase. |
| 3.5.1.15 Aspartoacylase. | 3.5.1.67 4-methyleneglutaminase. |

| | |
|---|---|
| 3.5.1.68 N-formylglutamate deformylase. | 3.5.3.9 Allantoate deiminase. |
| 3.5.1.69 Glycosphingolipid deacylase. | 3.5.3.10 D-arginase. |
| 3.5.1.70 Aculeacin-A deacylase. | 3.5.3.11 Agmatinase. |
| 3.5.1.71 N-feruloylglycine deacylase. | 3.5.3.12 Agmatine deiminase. |
| 3.5.1.72 D-benzoylarginine-4-nitroanilide amidase. | 3.5.3.13 Formimidoylglutamate deiminase. |
| 3.5.1.73 Carnitinamidase. | 3.5.3.14 Amidinoaspartase. |
| 3.5.1.74 Chenodeoxycholoyltaurine hydrolase. | 3.5.3.15 Protein-arginine deiminase. |
| 3.5.1.75 Urethanase. | 3.5.3.16 Methylguanidinase. |
| 3.5.1.76 Arylalkyl acylamidase. | 3.5.3.17 Guanidinopropionase. |
| 3.5.1.77 N-carbamoyl-D-amino acid hydrolase. | 3.5.3.18 Dimethylargininase. |
| 3.5.1.78 Glutathionylspermidine amidase. | 3.5.3.19 Ureidoglycolate hydrolase. |
| 3.5.1.79 Phthalyl amidase. | 3.5.3.20 Diguanidinobutanase. |
| 3.5.1.81 N-acyl-D-amino-acid deacylase. | 3.5.3.21 Methylenediurea deaminase. |
| 3.5.1.82 N-acyl-D-glutamate deacylase. | 3.5.3.22 Proclavaminate amidinohydrolase. |
| 3.5.1.83 N-acyl-D-aspartate deacylase. | 3.5.4.1 Cytosine deaminase. |

| | |
|---|---|
| 3.5.1.84 Biuret amidohydrolase. | 3.5.4.2 Adenine deaminase. |
| 3.5.1.85 (S)-N-acetyl-1-phenylethylamine hydrolase. | 3.5.4.3 Guanine deaminase. |
| 3.5.1.86 Mandelamide amidase. | 3.5.4.4 Adenosine deaminase. |
| 3.5.1.87 N-carbamoyl-L-amino-acid hydrolase. | 3.5.4.5 Cytidine deaminase. |
| 3.5.1.88 Peptide deformylase. | 3.5.4.6 AMP deaminase. |
| 3.5.1.89 N-acetylglucosaminylphosphatidylinositol deacetylase. | 3.5.4.7 ADP deaminase. |
| 3.5.1.90 Adenosylcobinamide hydrolase. | 3.5.4.8 Aminoimidazolase. |
| 3.5.2.1 Barbiturase. | 3.5.4.9 Methenyltetrahydrofolate cyclohydrolase. |
| 3.5.2.2 Dihydropyrimidinase. | 3.5.4.10 IMP cyclohydrolase. |
| 3.5.2.3 Dihydroorotase. | 3.5.4.11 Pterin deaminase. |
| 3.5.2.4 Carboxymethylhydantoinase. | 3.5.4.12 dCMP deaminase. |
| 3.5.2.5 Allantoinase. | 3.5.4.13 dCTP deaminase. |
| 3.5.2.6 Beta-lactamase. | 3.5.4.14 Deoxycytidine deaminase. |
| 3.5.2.7 Imidazolonepropionase. | 3.5.4.15 Guanosine deaminase. |
| 3.5.2.9 5-oxoprolinase (ATP-hydrolyzing). | 3.5.4.16 GTP cyclohydrolase I. |
| 3.5.2.10 Creatininase. | 3.5.4.17 Adenosine-phosphate deaminase. |
| 3.5.2.11 L-lysine-lactamase. | 3.5.4.18 ATP deaminase. |
| 3.5.2.12 6-aminohexanoate-cyclic-dimer hydrolase. | 3.5.4.19 Phosphoribosyl-AMP cyclohydrolase. |
| 3.5.2.13 2,5-dioxopiperazine hydrolase. | 3.5.4.20 Pyrithiamine deaminase. |
| 3.5.2.14 N-methylhydantoinase (ATP-hydrolyzing). | 3.5.4.21 Creatinine deaminase. |
| 3.5.2.15 Cyanuric acid amidohydrolase. | 3.5.4.22 1-pyrroline-4-hydroxy-2-carboxylate deaminase. |
| 3.5.2.16 Maleimide hydrolase. | 3.5.4.23 Blasticidin-S deaminase. |
| 3.5.2.17 Hydroxyisourate hydrolase. | 3.5.4.24 Sepiapterin deaminase. |
| 3.5.3.1 Arginase. | 3.5.4.25 GTP cyclohydrolase II. |
| 3.5.3.2 Guanidinoacetase. | 3.5.4.26 Diaminohydroxyphosphoribosylaminopyrimidine deaminase. |
| 3.5.3.3 Creatinase. | 3.5.4.27 Methenyltetrahydromethanopterin cyclohydrolase. |
| 3.5.3.4 Allantoicase. | 3.5.4.28 S-adenosylhomocysteine deaminase. |
| 3.5.3.5 Formimidoylaspartate deiminase. | 3.5.4.29 GTP cyclohydrolase IIa. |
| 3.5.3.6 Arginine deiminase. | 3.5.4.30 dCTP deaminase (dUMP-forming). |
| 3.5.3.7 Guanidinobutyrase. | 3.5.5.1 Nitrilase. |
| 3.5.3.8 Formimidoylglutamase. | 3.5.5.2 Ricinine nitrilase. |

| | |
|---|---|
| 3.5.5.4 Cyanoalanine nitrilase. | 3.6.1.40 Guanosine-5'-triphosphate,3'-diphosphate diphosphatase. |
| 3.5.5.5 Arylacetonitrilase. | 3.6.1.41 Bis(5'-nucleosyl)-tetraphosphatase (symmetrical). |
| 3.5.5.6 Bromoxynil nitrilase. | 3.6.1.42 Guanosine-diphosphatase. |
| 3.5.5.7 Aliphatic nitrilase. | 3.6.1.43 Dolichyldiphosphatase. |
| 3.5.5.8 Thiocyanate hydrolase. | 3.6.1.44 Oligosaccharide-diphosphodolichol diphosphatase. |
| 3.5.99.1 Riboflavinase. | 3.6.1.45 UDP-sugar diphosphatase. |
| 3.5.99.2 Thiaminase. | 3.6.1.52 Diphosphoinositol-polyphosphate diphosphatase. |
| 3.5.99.3 Hydroxydechloroatrazine ethylaminohydrolase. | 3.6.2.1 Adenylylsulfatase. |

| | |
|---|---|
| 3.5.99.4 N-isopropylammelide isopropylaminohydrolase. | 3.6.2.2 Phosphoadenylylsulfatase. |
| 3.5.99.5 2-aminomuconate deaminase. | 3.6.3.1 Phospholipid-translocating ATPase. |
| 3.5.99.6 Glucosamine-6-phosphate deaminase. | 3.6.3.2 Magnesium-importing ATPase. |
| 3.5.99.7 1-aminocyclopropane-1-carboxylate deaminase. | 3.6.3.3 Cadmium-exporting ATPase. |
| 3.6.1.1 Inorganic diphosphatase. | 3.6.3.4 Copper-exporting ATPase. |
| 3.6.1.2 Trimetaphosphatase. | 3.6.3.5 Zinc-exporting ATPase. |
| 3.6.1.3 Adenosinetriphosphatase. | 3.6.3.6 Proton-exporting ATPase. |
| 3.6.1.5 Apyrase. | 3.6.3.7 Sodium-exporting ATPase. |
| 3.6.1.6 Nucleoside-diphosphatase. | 3.6.3.8 Calcium-transporting ATPase. |
| 3.6.1.7 Acylphosphatase. | 3.6.3.9 Sodium/potassium-exchanging ATPase. |
| 3.6.1.8 ATP diphosphatase. | 3.6.3.10 Hydrogen/potassium-exchanging ATPase. |
| 3.6.1.9 Nucleotide diphosphatase. | 3.6.3.11 Chloride-transporting ATPase. |
| 3.6.1.10 Endopolyphosphatase. | 3.6.3.12 Potassium-transporting ATPase. |
| 3.6.1.11 Exopolyphosphatase. | 3.6.3.14 H(+)-transporting two-sector ATPase. |
| 3.6.1.12 dCTP diphosphatase. | 3.6.3.15 Sodium-transporting two-sector ATPase. |
| 3.6.1.13 ADP-ribose diphosphatase. | 3.6.3.16 Arsenite-transporting ATPase. |
| 3.6.1.14 Adenosine-tetraphosphatase. | 3.6.3.17 Monosaccharide-transporting ATPase. |
| 3.6.1.15 Nucleoside-triphosphatase. | 3.6.3.18 Oligosaccharide-transporting ATPase. |
| 3.6.1.16 CDP-glycerol diphosphatase. | 3.6.3.19 Maltose-transporting ATPase. |
| 3.6.1.17 Bis(5'-nucleosyl)-tetraphosphatase (asymmetrical). | 3.6.3.20 Glycerol-3-phosphate-transporting ATPase. |
| 3.6.1.18 FAD diphosphatase. | 3.6.3.21 Polar-amino-acid-transporting ATPase. |
| 3.6.1.19 Nucleoside-triphosphate diphosphatase. | 3.6.3.22 Nonpolar-amino-acid-transporting ATPase. |
| 3.6.1.20 5'-acylphosphoadenosine hydrolase. | 3.6.3.23 Oligopeptide-transporting ATPase. |
| 3.6.1.21 ADP-sugar diphosphatase. | 3.6.3.24 Nickel-transporting ATPase. |
| 3.6.1.22 NAD+ diphosphatase. | 3.6.3.25 Sulfate-transporting ATPase. |
| 3.6.1.23 dUTP diphosphatase. | 3.6.3.26 Nitrate-transporting ATPase. |
| 3.6.1.24 Nucleoside phosphoacylhydrolase. | 3.6.3.27 Phosphate-transporting ATPase. |
| 3.6.1.25 Triphosphatase. | 3.6.3.28 Phosphonate-transporting ATPase. |
| 3.6.1.26 CDP-diacylglycerol diphosphatase. | 3.6.3.29 Molybdate-transporting ATPase. |
| 3.6.1.27 Undecaprenyl-diphosphatase. | 3.6.3.30 Fe(3+)-transporting ATPase. |
| 3.6.1.28 Thiamine-triphosphatase. | 3.6.3.31 Polyamine-transporting ATPase. |
| 3.6.1.29 Bis(5'-adenosyl)-triphosphatase. | 3.6.3.32 Quaternary-amine-transporting ATPase. |
| 3.6.1.30 M(7)G(5')pppN diphosphatase. | 3.6.3.33 Vitamin B12-transporting ATPase. |
| 3.6.1.31 Phosphoribosyl-ATP diphosphatase. | 3.6.3.34 Iron-chelate-transporting ATPase. |
| 3.6.1.39 Thymidine-triphosphatase. | 3.6.3.35 Manganese-transporting ATPase. |

| | |
|---|---|
| 3.6.3.36 Taurine-transporting ATPase. | 3.8.1.1 Alkylhalidase. |
| 3.6.3.37 Guanine-transporting ATPase. | 3.8.1.2 (S)-2-haloacid dehalogenase. |
| 3.6.3.38 Capsular-polysaccharide-transporting ATPase. | 3.8.1.3 Haloacetate dehalogenase. |
| 3.6.3.39 Lipopolysaccharide-transporting ATPase. | 3.8.1.5 Haloalkane dehalogenase. |
| 3.6.3.40 Teichoic-acid-transporting ATPase. | 3.8.1.6 4-chlorobenzoate dehalogenase. |
| 3.6.3.41 Heme-transporting ATPase. | 3.8.1.7 4-chlorobenzoyl-CoA dehalogenase. |
| 3.6.3.42 Beta-glucan-transporting ATPase. | 3.8.1.8 Atrazine chlorohydrolase. |

| | |
|---|---|
| 3.6.3.43 Peptide-transporting ATPase. | 3.8.1.9 (R)-2-haloacid dehalogenase. |
| 3.6.3.44 Xenobiotic-transporting ATPase. | 3.8.1.10 2-haloacid dehalogenase (configuration-inverting). |
| 3.6.3.45 Steroid-transporting ATPase. | 3.8.1.11 2-haloacid dehalogenase (configuration-retaining). |
| 3.6.3.46 Cadmium-transporting ATPase. | 3.9.1.1 Phosphoamidase. |
| 3.6.3.47 Fatty-acyl-CoA-transporting ATPase. | 3.10.1.1 N-sulfoglucosamine sulfohydrolase. |
| 3.6.3.48 Alpha-factor-transporting ATPase. | 3.10.1.2 Cyclamate sulfohydrolase. |
| 3.6.3.49 Channel-conductance-controlling ATPase. | 3.11.1.1 Phosphonoacetaldehyde hydrolase. |
| 3.6.3.50 Protein-secreting ATPase. | 3.11.1.2 Phosphonoacetate hydrolase. |
| 3.6.3.51 Mitochondrial protein-transporting ATPase. | 3.12.1.1 Trithionate hydrolase. |
| 3.6.3.52 Chloroplast protein-transporting ATPase. | 3.13.1.1 UDP-sulfoquinovose synthase. |
| 3.6.3.53 Ag(+)-exporting ATPase. | |
| 3.6.4.1 Myosin ATPase. | ENZYME: 4.-.-.- |
| 3.6.4.2 Dynein ATPase. | 4.1.1.1 Pyruvate decarboxylase. |
| 3.6.4.3 Microtubule-severing ATPase. | 4.1.1.2 Oxalate decarboxylase. |
| 3.6.4.4 Plus-end-directed kinesin ATPase. | 4.1.1.3 Oxaloacetate decarboxylase. |
| 3.6.4.5 Minus-end-directed kinesin ATPase. | 4.1.1.4 Acetoacetate decarboxylase. |
| 3.6.4.6 Vesicle-fusing ATPase. | 4.1.1.5 Acetolactate decarboxylase. |
| 3.6.4.7 Peroxisome-assembly ATPase. | 4.1.1.6 Aconitate decarboxylase. |
| 3.6.4.8 Proteasome ATPase. | 4.1.1.7 Benzoylformate decarboxylase. |
| 3.6.4.9 Chaperonin ATPase. | 4.1.1.8 Oxalyl-CoA decarboxylase. |
| 3.6.4.10 Non-chaperonin molecular chaperone ATPase. | 4.1.1.9 Malonyl-CoA decarboxylase. |
| 3.6.4.11 Nucleoplasmin ATPase. | 4.1.1.11 Aspartate 1-decarboxylase. |
| 3.6.5.1 Heterotrimeric G-protein GTPase. | 4.1.1.12 Aspartate 4-decarboxylase. |
| 3.6.5.2 Small monomeric GTPase. | 4.1.1.14 Valine decarboxylase. |
| 3.6.5.3 Protein-synthesizing GTPase. | 4.1.1.15 Glutamate decarboxylase. |
| 3.6.5.4 Signal-recognition-particle GTPase. | 4.1.1.16 Hydroxyglutamate decarboxylase. |
| 3.6.5.5 Dynamin GTPase. | 4.1.1.17 Ornithine decarboxylase. |
| 3.6.5.6 Tubulin GTPase. | 4.1.1.18 Lysine decarboxylase. |
| 3.7.1.1 Oxaloacetase. | 4.1.1.19 Arginine decarboxylase. |
| 3.7.1.2 Fumarylacetoacetase. | 4.1.1.20 Diaminopimelate decarboxylase. |
| 3.7.1.3 Kynureninase. | 4.1.1.21 Phosphoribosylaminoimidazole carboxylase. |
| 3.7.1.4 Phloretin hydrolase. | 4.1.1.22 Histidine decarboxylase. |
| 3.7.1.5 Acylpyruvate hydrolase. | 4.1.1.23 Orotidine-5'-phosphate decarboxylase. |
| 3.7.1.6 Acetylpyruvate hydrolase. | 4.1.1.24 Aminobenzoate decarboxylase. |
| 3.7.1.7 Beta-diketone hydrolase. | 4.1.1.25 Tyrosine decarboxylase. |
| 3.7.1.8 2,6-dioxo-6-phenylhexa-3-enoate hydrolase. | 4.1.1.28 Aromatic-L-amino-acid decarboxylase. |
| 3.7.1.9 2-hydroxymuconate-semialdehyde hydrolase. | 4.1.1.29 Sulfinoalanine decarboxylase. |
| 3.7.1.10 Cyclohexane-1,3-dione hydrolase. | 4.1.1.30 Pantothenoylcysteine decarboxylase. |

| | |
|---|---|
| 4.1.1.31 Phosphoenolpyruvate carboxylase. | 4.1.1.74 Indolepyruvate decarboxylase. |
| 4.1.1.32 Phosphoenolpyruvate carboxykinase (GTP). | 4.1.1.75 5-guanidino-2-oxopentanoate decarboxylase. |
| 4.1.1.33 Diphosphomevalonate decarboxylase. | 4.1.1.76 Arylmalonate decarboxylase. |
| 4.1.1.34 Dehydro-L-gulonate decarboxylase. | 4.1.1.77 4-oxalocrotonate decarboxylase. |
| 4.1.1.35 UDP-glucuronate decarboxylase. | 4.1.1.78 Acetylenedicarboxylate decarboxylase. |

| | |
|---|---|
| 4.1.1.36 Phosphopantothenoylcysteine decarboxylase. | 4.1.1.79 Sulfopyruvate decarboxylase. |
| 4.1.1.37 Uroporphyrinogen decarboxylase. | 4.1.1.80 4-hydroxyphenylpyruvate decarboxylase. |
| 4.1.1.38 Phosphoenolpyruvate carboxykinase (diphosphate). | 4.1.1.81 Threonine-phosphate decarboxylase. |
| 4.1.1.39 Ribulose-bisphosphate carboxylase. | 4.1.2.2 Ketotetrose-phosphate aldolase. |
| 4.1.1.40 Hydroxypyruvate decarboxylase. | 4.1.2.4 Deoxyribose-phosphate aldolase. |
| 4.1.1.41 Methylmalonyl-CoA decarboxylase. | 4.1.2.5 Threonine aldolase. |
| 4.1.1.42 Carnitine decarboxylase. | 4.1.2.9 Phosphoketolase. |
| 4.1.1.43 Phenylpyruvate decarboxylase. | 4.1.2.10 Mandelonitrile lyase. |
| 4.1.1.44 4-carboxymuconolactone decarboxylase. | 4.1.2.11 Hydroxymandelonitrile lyase. |
| 4.1.1.45 Aminocarboxymuconate-semialdehyde decarboxylase. | 4.1.2.12 2-dehydropantoate aldolase. |
| 4.1.1.46 O-pyrocatechuate decarboxylase. | 4.1.2.13 Fructose-bisphosphate aldolase. |
| 4.1.1.47 Tartronate-semialdehyde synthase. | 4.1.2.14 2-dehydro-3-deoxy-phosphogluconate aldolase. |
| 4.1.1.48 Indole-3-glycerol-phosphate synthase. | 4.1.2.17 L-fuculose-phosphate aldolase. |
| 4.1.1.49 Phosphoenolpyruvate carboxykinase (ATP). | 4.1.2.18 2-dehydro-3-deoxy-L-pentonate aldolase. |
| 4.1.1.50 Adenosylmethionine decarboxylase. | 4.1.2.19 Rhamnulose-1-phosphate aldolase. |
| 4.1.1.51 3-hydroxy-2-methylpyridine-4,5-dicarboxylate 4-decarboxylase. | 4.1.2.20 2-dehydro-3-deoxyglucarate aldolase. |
| 4.1.1.52 6-methylsalicylate decarboxylase. | 4.1.2.21 2-dehydro-3-deoxy-6-phosphogalactonate aldolase. |
| 4.1.1.53 Phenylalanine decarboxylase. | 4.1.2.22 Fructose-6-phosphate phosphoketolase. |
| 4.1.1.54 Dihydroxyfumarate decarboxylase. | 4.1.2.23 3-deoxy-D-manno-octulosonate aldolase. |
| 4.1.1.55 4,5-dihydroxyphthalate decarboxylase. | 4.1.2.24 Dimethylaniline-N-oxide aldolase. |
| 4.1.1.56 3-oxolaurate decarboxylase. | 4.1.2.25 Dihydroneopterin aldolase. |
| 4.1.1.57 Methionine decarboxylase. | 4.1.2.26 Phenylserine aldolase. |
| 4.1.1.58 Orsellinate decarboxylase. | 4.1.2.27 Sphinganine-1-phosphate aldolase. |
| 4.1.1.59 Gallate decarboxylase. | 4.1.2.28 2-dehydro-3-deoxy-D-pentonate aldolase. |
| 4.1.1.60 Stipitatonate decarboxylase. | 4.1.2.29 5-dehydro-2-deoxyphosphogluconate aldolase. |
| 4.1.1.61 4-hydroxybenzoate decarboxylase. | 4.1.2.30 17-alpha-hydroxyprogesterone aldolase. |
| 4.1.1.62 Gentisate decarboxylase. | 4.1.2.32 Trimethylamine-oxide aldolase. |
| 4.1.1.63 Protocatechuate decarboxylase. | 4.1.2.33 Fucosterol-epoxide lyase. |
| 4.1.1.64 2,2-dialkylglycine decarboxylase (pyruvate). | 4.1.2.34 4-(2-carboxyphenyl)-2-oxobut-3-enoate aldolase. |
| 4.1.1.65 Phosphatidylserine decarboxylase. | 4.1.2.35 Propioin synthase. |
| 4.1.1.66 Uracil-5-carboxylate decarboxylase. | 4.1.2.36 Lactate aldolase. |
| 4.1.1.67 UDP-galacturonate decarboxylase. | 4.1.2.37 Acetone-cyanohydrin lyase. |
| 4.1.1.68 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase. | 4.1.2.38 Benzoin aldolase. |
| 4.1.1.69 3,4-dihydroxyphthalate decarboxylase. | 4.1.2.39 Hydroxynitrilase. |
| 4.1.1.70 Glutaconyl-CoA decarboxylase. | 4.1.2.40 Tagatose-bisphosphate aldolase. |
| 4.1.1.71 2-oxoglutarate decarboxylase. | 4.1.2.41 Vanillin synthase. |
| 4.1.1.72 Branched-chain-2-oxoacid decarboxylase. | 4.1.3.1 Isocitrate lyase. |
| 4.1.1.73 Tartrate decarboxylase. | 4.1.3.3 N-acetylneuraminate lyase. |

| | |
|---|---|
| 4.1.3.4 Hydroxymethylglutaryl-CoA lyase. | 4.2.1.32 L(+)-tartrate dehydratase. |
| 4.1.3.6 Citrate (pro-3S)-lyase. | 4.2.1.33 3-isopropylmalate dehydratase. |

| | |
|---|---|
| 4.1.3.13 Oxalomalate lyase. | 4.2.1.34 (S)-2-methylmalate dehydratase. |
| 4.1.3.14 3-hydroxyaspartate aldolase. | 4.2.1.35 (R)-2-methylmalate dehydratase. |
| 4.1.3.16 4-hydroxy-2-oxoglutarate aldolase. | 4.2.1.36 Homoaconitate hydratase. |
| 4.1.3.17 4-hydroxy-4-methyl-2-oxoglutarate aldolase. | 4.2.1.39 Gluconate dehydratase. |
| 4.1.3.22 Citramalate lyase. | 4.2.1.40 Glucarate dehydratase. |
| 4.1.3.24 Malyl-CoA lyase. | 4.2.1.41 5-dehydro-4-deoxyglucarate dehydratase. |
| 4.1.3.25 Citramalyl-CoA lyase. | 4.2.1.42 Galactarate dehydratase. |
| 4.1.3.26 3-hydroxy-3-isohexenylglutaryl-CoA lyase. | 4.2.1.43 2-dehydro-3-deoxy-L-arabinonate dehydratase. |
| 4.1.3.27 Anthranilate synthase. | 4.2.1.44 Myo-inosose-2 dehydratase. |
| 4.1.3.30 Methylisocitrate lyase. | 4.2.1.45 CDP-glucose 4,6-dehydratase. |
| 4.1.3.32 2,3-dimethylmalate lyase. | 4.2.1.46 dTDP-glucose 4,6-dehydratase. |
| 4.1.3.34 Citryl-CoA lyase. | 4.2.1.47 GDP-mannose 4,6-dehydratase. |
| 4.1.3.35 (1-hydroxycyclohexan-1-yl)acetyl-CoA lyase. | 4.2.1.48 D-glutamate cyclase. |
| 4.1.3.36 Naphthoate synthase. | 4.2.1.49 Urocanate hydratase. |
| 4.1.3.38 Aminodeoxychorismate lyase. | 4.2.1.50 Pyrazolylalanine synthase. |
| 4.1.99.1 Tryptophanase. | 4.2.1.51 Prephenate dehydratase. |
| 4.1.99.2 Tyrosine phenol-lyase. | 4.2.1.52 Dihydrodipicolinate synthase. |
| 4.1.99.3 Deoxyribodipyrimidine photo-lyase. | 4.2.1.53 Oleate hydratase. |
| 4.1.99.5 Octadecanal decarbonylase. | 4.2.1.54 Lactoyl-CoA dehydratase. |
| 4.1.99.11 Benzylsuccinate synthase. | 4.2.1.55 3-hydroxybutyryl-CoA dehydratase. |
| 4.2.1.1 Carbonate dehydratase. | 4.2.1.56 Itaconyl-CoA hydratase. |
| 4.2.1.2 Fumarate hydratase. | 4.2.1.57 Isohexenylglutaconyl-CoA hydratase. |
| 4.2.1.3 Aconitate hydratase. | 4.2.1.58 Crotonoyl-[acyl-carrier-protein] hydratase. |
| 4.2.1.4 Citrate dehydratase. | 4.2.1.59 3-hydroxyoctanoyl-[acyl-carrier-protein] dehydratase. |
| 4.2.1.5 Arabinonate dehydratase. | 4.2.1.60 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase. |
| 4.2.1.6 Galactonate dehydratase. | 4.2.1.61 3-hydroxypalmitoyl-[acyl-carrier-protein] dehydratase. |
| 4.2.1.7 Altronate dehydratase. | 4.2.1.62 5-alpha-hydroxysteroid dehydratase. |
| 4.2.1.8 Mannonate dehydratase. | 4.2.1.65 3-cyanoalanine hydratase. |
| 4.2.1.9 Dihydroxy-acid dehydratase. | 4.2.1.66 Cyanide hydratase. |
| 4.2.1.10 3-dehydroquinate dehydratase. | 4.2.1.67 D-fuconate dehydratase. |
| 4.2.1.11 Phosphopyruvate hydratase. | 4.2.1.68 L-fuconate dehydratase. |
| 4.2.1.12 Phosphogluconate dehydratase. | 4.2.1.69 Cyanamide hydratase. |
| 4.2.1.17 Enoyl-CoA hydratase. | 4.2.1.70 Pseudouridylate synthase. |
| 4.2.1.18 Methylglutaconyl-CoA hydratase. | 4.2.1.73 Protoaphin-aglucone dehydratase (cyclizing). |
| 4.2.1.19 Imidazoleglycerol-phosphate dehydratase. | 4.2.1.74 Long-chain-enoyl-CoA hydratase. |
| 4.2.1.20 Tryptophan synthase. | 4.2.1.75 Uroporphyrinogen-III synthase. |
| 4.2.1.22 Cystathionine beta-synthase. | 4.2.1.76 UDP-glucose 4,6-dehydratase. |
| 4.2.1.24 Porphobilinogen synthase. | 4.2.1.77 Trans-L-3-hydroxyproline dehydratase. |
| 4.2.1.25 L-arabinonate dehydratase. | 4.2.1.78 (S)-norcoclaurine synthase. |
| 4.2.1.27 Acetylenecarboxylate hydratase. | 4.2.1.79 2-methylcitrate dehydratase. |
| 4.2.1.28 Propanediol dehydratase. | 4.2.1.80 2-oxopent-4-enoate hydratase. |
| 4.2.1.30 Glycerol dehydratase. | 4.2.1.81 D(-)-tartrate dehydratase. |
| 4.2.1.31 Maleate hydratase. | 4.2.1.82 Xylonate dehydratase. |

| | |
|---|---|
| 4.2.1.83 4-oxalmesaconate hydratase. | 4.2.3.6 Trichodiene synthase. |
| 4.2.1.84 Nitrile hydratase. | 4.2.3.7 Pentalenene synthase. |
| 4.2.1.85 Dimethylmaleate hydratase. | 4.2.3.8 Casbene synthase. |
| 4.2.1.86 16-dehydroprogesterone hydratase. | 4.2.3.9 Aristolochene synthase. |
| 4.2.1.87 Octopamine dehydratase. | 4.2.3.10 (-)-endo-fenchol synthase. |
| 4.2.1.88 Synephrine dehydratase. | 4.2.3.11 Sabinene-hydrate synthase. |
| 4.2.1.89 Carnitine dehydratase. | 4.2.3.12 6-pyruvoyltetrahydropterin synthase. |
| 4.2.1.90 L-rhamnonate dehydratase. | 4.2.3.13 (+)-delta-cadinene synthase. |
| 4.2.1.91 Carboxycyclohexadienyl dehydratase. | 4.2.3.14 Pinene synthase. |
| 4.2.1.92 Hydroperoxide dehydratase. | 4.2.3.15 Myrcene synthase. |
| 4.2.1.93 ATP-dependent NAD(P)H-hydrate dehydratase. | 4.2.3.16 (4S)-limonene synthase. |
| 4.2.1.94 Scytalone dehydratase. | 4.2.3.17 Taxadiene synthase. |
| 4.2.1.95 Kievitone hydratase. | 4.2.3.18 Abietadiene synthase. |
| 4.2.1.96 4a-hydroxytetrahydrobiopterin dehydratase. | 4.2.3.19 Ent-kaurene synthase. |
| 4.2.1.97 Phaseollidin hydratase. | 4.2.3.20 (+)-limonene synthase. |
| 4.2.1.98 16-alpha-hydroxyprogesterone dehydratase. | 4.2.3.21 Vetispiradiene synthase. |
| 4.2.1.99 2-methylisocitrate dehydratase. | 4.2.99.12 Carboxymethyloxysuccinate lyase. |
| 4.2.1.100 Cyclohexa-1,5-dienecarbonyl-CoA hydratase. | 4.2.99.18 DNA-(apurinic or apyrimidinic site) lyase. |
| 4.2.1.101 Trans-feruloyl-CoA hydratase. | 4.2.99.19 2-hydroxypropyl-CoM lyase. |
| 4.2.1.103 Cyclohexyl-isocyanide hydratase. | 4.3.1.1 Aspartate ammonia-lyase. |
| 4.2.1.104 Cyanate hydratase. | 4.3.1.2 Methylaspartate ammonia-lyase. |
| 4.2.2.1 Hyaluronate lyase. | 4.3.1.3 Histidine ammonia-lyase. |
| 4.2.2.2 Pectate lyase. | 4.3.1.4 Formimidoyltetrahydrofolate cyclodeaminase. |
| 4.2.2.3 Poly(beta-D-mannuronate) lyase. | 4.3.1.5 Phenylalanine ammonia-lyase. |
| 4.2.2.4 Chondroitin ABC lyase. | 4.3.1.6 Beta-alanyl-CoA ammonia-lyase. |
| 4.2.2.5 Chondroitin AC lyase. | 4.3.1.7 Ethanolamine ammonia-lyase. |
| 4.2.2.6 Oligogalacturonide lyase. | 4.3.1.9 Glucosaminate ammonia-lyase. |
| 4.2.2.7 Heparin lyase. | 4.3.1.10 Serine-sulfate ammonia-lyase. |
| 4.2.2.8 Heparin-sulfate lyase. | 4.3.1.11 Dihydroxyphenylalanine ammonia-lyase. |
| 4.2.2.9 Pectate disaccharide-lyase. | 4.3.1.12 Ornithine cyclodeaminase. |
| 4.2.2.10 Pectin lyase. | 4.3.1.13 Carbamoyl-serine ammonia-lyase. |
| 4.2.2.11 Poly(alpha-L-guluronate) lyase. | 4.3.1.14 3-aminobutyryl-CoA ammonia-lyase. |
| 4.2.2.12 Xanthan lyase. | 4.3.1.15 Diaminopropionate ammonia-lyase. |
| 4.2.2.13 Exo-(1->4)-alpha-D-glucan lyase. | 4.3.1.16 Threo-3-hydroxyaspartate ammonia-lyase. |
| 4.2.2.14 Glucuronan lyase. | 4.3.1.17 L-serine ammonia-lyase. |
| 4.2.2.15 Anhydrosialidase. | 4.3.1.18 D-serine ammonia-lyase. |
| 4.2.2.16 Levan fructotransferase (DFA-IV-forming). | 4.3.1.19 Threonine ammonia-lyase. |
| 4.2.2.17 Inulin fructotransferase (DFA-I-forming). | 4.3.1.20 Erythro-3-hydroxyaspartate ammonia-lyase. |
| 4.2.2.18 Inulin fructotransferase (DFA-III-forming). | 4.3.2.1 Argininosuccinate lyase. |
| 4.2.3.1 Threonine synthase. | 4.3.2.2 Adenylosuccinate lyase. |
| 4.2.3.2 Ethanolamine-phosphate phospho-lyase. | 4.3.2.3 Ureidoglycolate lyase. |
| 4.2.3.3 Methylglyoxal synthase. | 4.3.2.4 Purine imidazole-ring cyclase. |
| 4.2.3.4 3-dehydroquinate synthase. | 4.3.2.5 Peptidylamidoglycolate lyase. |
| 4.2.3.5 Chorismate synthase. | 4.3.3.1 3-ketovalidoxylamine C-N-lyase. |

| | |
|---|---|
| 4.3.3.2 Strictosidine synthase. | 5.1.1.7 Diaminopimelate epimerase. |
| 4.3.3.3 Deacetylisoipecoside synthase. | 5.1.1.8 4-hydroxyproline epimerase. |
| 4.3.3.4 Deacetylipecoside synthase. | 5.1.1.9 Arginine racemase. |
| 4.4.1.1 Cystathionine gamma-lyase. | 5.1.1.10 Amino-acid racemase. |
| 4.4.1.2 Homocysteine desulfhydrase. | 5.1.1.11 Phenylalanine racemase (ATP-hydrolyzing). |
| 4.4.1.3 Dimethylpropiothetin dethiomethylase. | 5.1.1.12 Ornithine racemase. |
| 4.4.1.4 Alliin lyase. | 5.1.1.13 Aspartate racemase. |
| 4.4.1.5 Lactoylglutathione lyase. | 5.1.1.14 Nocardicin-A epimerase. |
| 4.4.1.6 S-alkylcysteine lyase. | 5.1.1.15 2-aminohexano-6-lactam racemase. |
| 4.4.1.8 Cystathionine beta-lyase. | 5.1.1.16 Protein-serine epimerase. |
| 4.4.1.9 L-3-cyanoalanine synthase. | 5.1.1.17 Isopenicillin-N epimerase. |
| 4.4.1.10 Cysteine lyase. | 5.1.2.1 Lactate racemase. |
| 4.4.1.11 Methionine gamma-lyase. | 5.1.2.2 Mandelate racemase. |
| 4.4.1.13 Cysteine-S-conjugate beta-lyase. | 5.1.2.3 3-hydroxybutyryl-CoA epimerase. |
| 4.4.1.14 1-aminocyclopropane-1-carboxylate synthase. | 5.1.2.4 Acetoin racemase. |
| 4.4.1.15 D-cysteine desulfhydrase. | 5.1.2.5 Tartrate epimerase. |
| 4.4.1.16 Selenocysteine lyase. | 5.1.2.6 Isocitrate epimerase. |
| 4.4.1.17 Holocytochrome-c synthase. | 5.1.3.1 Ribulose-phosphate 3-epimerase. |
| 4.4.1.19 Phosphosulfolactate synthase. | 5.1.3.2 UDP-glucose 4-epimerase. |
| 4.4.1.20 Leukotriene-C(4) synthase. | 5.1.3.3 Aldose 1-epimerase. |
| 4.5.1.1 DDT-dehydrochlorinase. | 5.1.3.4 L-ribulose-phosphate 4-epimerase. |
| 4.5.1.2 3-chloro-D-alanine dehydrochlorinase. | 5.1.3.5 UDP-arabinose 4-epimerase. |
| 4.5.1.3 Dichloromethane dehalogenase. | 5.1.3.6 UDP-glucuronate 4-epimerase. |
| 4.5.1.4 L-2-amino-4-chloropent-4-enoate dehydrochlorinase. | 5.1.3.7 UDP-N-acetylglucosamine 4-epimerase. |
| 4.5.1.5 S-carboxymethylcysteine synthase. | 5.1.3.8 N-acylglucosamine 2-epimerase. |
| 4.6.1.1 Adenylate cyclase. | 5.1.3.9 N-acylglucosamine-6-phosphate 2-epimerase. |
| 4.6.1.2 Guanylate cyclase. | 5.1.3.10 CDP-abequose epimerase. |
| 4.6.1.6 Cytidylate cyclase. | 5.1.3.11 Cellobiose epimerase. |
| 4.6.1.12 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase. | 5.1.3.12 UDP-glucuronate 5'-epimerase. |
| 4.6.1.13 Phosphatidylinositol diacylglycerol-lyase. | 5.1.3.13 dTDP-4-dehydrorhamnose 3,5-epimerase. |
| 4.6.1.14 Glycosylphosphatidylinositol diacylglycerol-lyase. | 5.1.3.14 UDP-N-acetylglucosamine 2-epimerase. |
| 4.6.1.15 FAD-AMP lyase (cyclizing). | 5.1.3.15 Glucose-6-phosphate 1-epimerase. |
| 4.99.1.1 Ferrochelatase. | 5.1.3.16 UDP-glucosamine 4-epimerase. |
| 4.99.1.2 Alkylmercury lyase. | 5.1.3.17 Heparosan-N-sulfate-glucuronate 5-epimerase. |
| 4.99.1.3 Sirohydrochlorin cobaltochelatase. | 5.1.3.18 GDP-mannose 3,5-epimerase. |
| 4.99.1.4 Sirohydrochlorin ferrochelatase. | 5.1.3.19 Chondroitin-glucuronate 5-epimerase. |
| 4.99.1.5 Aliphatic aldoxime dehydratase. | 5.1.3.20 ADP-glyceromanno-heptose 6-epimerase. |
| 4.99.1.6 Indoleacetaldoxime dehydratase. | 5.1.3.21 Maltose epimerase. |
| | 5.1.99.1 Methylmalonyl-CoA epimerase. |
| ENZYME: 5.-.-.- | 5.1.99.2 16-hydroxysteroid epimerase. |
| 5.1.1.1 Alanine racemase. | 5.1.99.3 Allantoin racemase. |
| 5.1.1.2 Methionine racemase. | 5.1.99.4 Alpha-methylacyl-CoA racemase. |
| 5.1.1.3 Glutamate racemase. | 5.2.1.1 Maleate isomerase. |

| | |
|---|---|
| 5.1.1.4 Proline racemase. | 5.2.1.2 Maleylacetoacetate isomerase. |
| 5.1.1.5 Lysine racemase. | 5.2.1.3 Retinal isomerase. |
| 5.1.1.6 Threonine racemase. | 5.2.1.4 Maleylpyruvate isomerase. |

| | |
|---|---|
| 5.2.1.5 Linoleate isomerase. | 5.3.4.1 Protein disulfide-isomerase. |
| 5.2.1.6 Furylfuramide isomerase. | 5.3.99.2 Prostaglandin-D synthase. |
| 5.2.1.7 Retinol isomerase. | 5.3.99.3 Prostaglandin-E synthase. |
| 5.2.1.8 Peptidylprolyl isomerase. | 5.3.99.4 Prostaglandin-I synthase. |
| 5.2.1.9 Farnesol 2-isomerase. | 5.3.99.5 Thromboxane-A synthase. |
| 5.2.1.10 2-chloro-4-carboxymethylenebut-2-en-1,4-olide isomerase. | 5.3.99.6 Allene-oxide cyclase. |
| 5.2.1.11 4-hydroxyphenylacetaldehyde-oxime isomerase. | 5.3.99.7 Styrene-oxide isomerase. |
| 5.3.1.1 Triose-phosphate isomerase. | 5.4.1.1 Lysolecithin acylmutase. |
| 5.3.1.3 Arabinose isomerase. | 5.4.1.2 Precorrin-8X methylmutase. |
| 5.3.1.4 L-arabinose isomerase. | 5.4.2.1 Phosphoglycerate mutase. |
| 5.3.1.5 Xylose isomerase. | 5.4.2.2 Phosphoglucomutase. |
| 5.3.1.6 Ribose-5-phosphate isomerase. | 5.4.2.3 Phosphoacetylglucosamine mutase. |
| 5.3.1.7 Mannose isomerase. | 5.4.2.4 Bisphosphoglycerate mutase. |
| 5.3.1.8 Mannose-6-phosphate isomerase. | 5.4.2.5 Phosphoglucomutase (glucose-cofactor). |
| 5.3.1.9 Glucose-6-phosphate isomerase. | 5.4.2.6 Beta-phosphoglucomutase. |
| 5.3.1.12 Glucuronate isomerase. | 5.4.2.7 Phosphopentomutase. |
| 5.3.1.13 Arabinose-5-phosphate isomerase. | 5.4.2.8 Phosphomannomutase. |
| 5.3.1.14 L-rhamnose isomerase. | 5.4.2.9 Phosphoenolpyruvate mutase. |
| 5.3.1.15 D-lyxose ketol-isomerase. | 5.4.2.10 Phosphoglucosamine mutase. |
| 5.3.1.16 1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4-carboxamide isomerase. | 5.4.3.2 Lysine 2,3-aminomutase. |
| 5.3.1.17 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase. | 5.4.3.3 Beta-lysine 5,6-aminomutase. |
| 5.3.1.20 Ribose isomerase. | 5.4.3.4 D-lysine 5,6-aminomutase. |
| 5.3.1.21 Corticosteroid side-chain-isomerase. | 5.4.3.5 D-ornithine 4,5-aminomutase. |
| 5.3.1.22 Hydroxypyruvate isomerase. | 5.4.3.6 Tyrosine 2,3-aminomutase. |
| 5.3.1.23 S-methyl-5-thioribose-1-phosphate isomerase. | 5.4.3.7 Leucine 2,3-aminomutase. |
| 5.3.1.24 Phosphoribosylanthranilate isomerase. | 5.4.3.8 Glutamate-1-semialdehyde 2,1-aminomutase. |
| 5.3.1.25 L-fucose isomerase. | 5.4.4.1 (Hydroxyamino)benzene mutase. |
| 5.3.1.26 Galactose-6-phosphate isomerase. | 5.4.4.2 Isochorismate synthase. |
| 5.3.2.1 Phenylpyruvate tautomerase. | 5.4.4.3 3-(hydroxyamino)phenol mutase. |
| 5.3.2.2 Oxaloacetate tautomerase. | 5.4.99.1 Methylaspartate mutase. |
| 5.3.3.1 Steroid delta-isomerase. | 5.4.99.2 Methylmalonyl-CoA mutase. |
| 5.3.3.2 Isopentenyl-diphosphate delta-isomerase. | 5.4.99.3 2-acetolactate mutase. |
| 5.3.3.3 Vinylacetyl-CoA delta-isomerase. | 5.4.99.4 2-methyleneglutarate mutase. |
| 5.3.3.4 Muconolactone delta-isomerase. | 5.4.99.5 Chorismate mutase. |
| 5.3.3.5 Cholestenol delta-isomerase. | 5.4.99.7 Lanosterol synthase. |
| 5.3.3.6 Methylitaconate delta-isomerase. | 5.4.99.8 Cycloartenol synthase. |
| 5.3.3.7 Aconitate delta-isomerase. | 5.4.99.9 UDP-galactopyranose mutase. |
| 5.3.3.8 Dodecenoyl-CoA delta-isomerase. | 5.4.99.11 Isomaltulose synthase. |
| 5.3.3.9 Prostaglandin-A(1) delta-isomerase. | 5.4.99.12 tRNA-pseudouridine synthase I. |
| 5.3.3.10 5-carboxymethyl-2-hydroxymuconate delta-isomerase. | 5.4.99.13 Isobutyryl-CoA mutase. |

| | |
|---|---|
| 5.3.3.11 Isopiperitenone delta-isomerase. | 5.4.99.14 4-carboxymethyl-4-methylbutenolide mutase. |
| 5.3.3.12 Dopachrome isomerase. | 5.4.99.15 (1->4)-alpha-D-glucan 1-alpha-D-glucosylmutase. |
| 5.3.3.13 Polyenoic fatty acid isomerase. | 5.4.99.16 Maltose alpha-D-glucosyltransferase. |

| | |
|---|---|
| 5.4.99.17 Squalene--hopene cyclase. | 6.2.1.5 Succinate--CoA ligase (ADP-forming). |
| 5.5.1.1 Muconate cycloisomerase. | 6.2.1.6 Glutarate--CoA ligase. |
| 5.5.1.2 3-carboxy-cis,cis-muconate cycloisomerase. | 6.2.1.7 Cholate--CoA ligase. |
| 5.5.1.3 Tetrahydroxypteridine cycloisomerase. | 6.2.1.8 Oxalate--CoA ligase. |
| 5.5.1.4 Inositol-3-phosphate synthase. | 6.2.1.9 Malate--CoA ligase. |
| 5.5.1.5 Carboxy-cis,cis-muconate cyclase. | 6.2.1.10 Acid--CoA ligase (GDP-forming). |
| 5.5.1.6 Chalcone isomerase. | 6.2.1.11 Biotin--CoA ligase. |
| 5.5.1.7 Chloromuconate cycloisomerase. | 6.2.1.12 4-coumarate--CoA ligase. |
| 5.5.1.8 Geranyl-diphosphate cyclase. | 6.2.1.13 Acetate--CoA ligase (ADP-forming). |
| 5.5.1.9 Cycloeucalenol cycloisomerase. | 6.2.1.14 6-carboxyhexanoate--CoA ligase. |
| 5.5.1.10 Alpha-pinene-oxide decyclase. | 6.2.1.15 Arachidonate--CoA ligase. |
| 5.5.1.11 Dichloromuconate cycloisomerase. | 6.2.1.16 Acetoacetate--CoA ligase. |
| 5.5.1.12 Copalyl diphosphate synthase. | 6.2.1.17 Propionate--CoA ligase. |
| 5.5.1.13 Ent-copalyl diphosphate synthase. | 6.2.1.18 Citrate--CoA ligase. |
| 5.99.1.1 Thiocyanate isomerase. | 6.2.1.19 Long-chain-fatty-acid--luciferin-component ligase. |
| 5.99.1.2 DNA topoisomerase. | 6.2.1.20 Long-chain-fatty-acid--[acyl-carrier-protein] ligase. |
| 5.99.1.3 DNA topoisomerase (ATP-hydrolyzing). | 6.2.1.22 [Citrate (pro-3S)-lyase] ligase. |
| ENZYME: 6.-.-.- | 6.2.1.23 Dicarboxylate--CoA ligase. |
| 6.1.1.1 Tyrosine--tRNA ligase. | 6.2.1.24 Phytanate--CoA ligase. |
| 6.1.1.2 Tryptophan--tRNA ligase. | 6.2.1.25 Benzoate--CoA ligase. |
| 6.1.1.3 Threonine--tRNA ligase. | 6.2.1.26 O-succinylbenzoate--CoA ligase. |
| 6.1.1.4 Leucine--tRNA ligase. | 6.2.1.27 4-hydroxybenzoate--CoA ligase. |
| 6.1.1.5 Isoleucine--tRNA ligase. | 6.2.1.28 3-alpha,7-alpha-dihydroxy-5-beta-cholestanate--CoA ligase. |
| 6.1.1.6 Lysine--tRNA ligase. | 6.2.1.29 3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholestanate--CoA ligase. |
| 6.1.1.7 Alanine--tRNA ligase. | 6.2.1.30 Phenylacetate--CoA ligase. |
| 6.1.1.9 Valine--tRNA ligase. | 6.2.1.31 2-furoate--CoA ligase. |
| 6.1.1.10 Methionine--tRNA ligase. | 6.2.1.32 Anthranilate--CoA ligase. |
| 6.1.1.11 Serine--tRNA ligase. | 6.2.1.33 4-chlorobenzoate--CoA ligase. |
| 6.1.1.12 Aspartate--tRNA ligase. | 6.2.1.34 Trans-feruloyl--CoA synthase. |
| 6.1.1.13 D-alanine--poly(phosphoribitol) ligase. | 6.3.1.1 Aspartate--ammonia ligase. |
| 6.1.1.14 Glycine--tRNA ligase. | 6.3.1.2 Glutamate--ammonia ligase. |
| 6.1.1.15 Proline--tRNA ligase. | 6.3.1.4 Aspartate--ammonia ligase (ADP-forming). |
| 6.1.1.16 Cysteine--tRNA ligase. | 6.3.1.5 NAD(+) synthase. |
| 6.1.1.17 Glutamate--tRNA ligase. | 6.3.1.6 Glutamate--ethylamine ligase. |
| 6.1.1.18 Glutamine--tRNA ligase. | 6.3.1.7 4-methyleneglutamate--ammonia ligase. |
| 6.1.1.19 Arginine--tRNA ligase. | 6.3.1.8 Glutathionylspermidine synthase. |
| 6.1.1.20 Phenylalanine--tRNA ligase. | 6.3.1.9 Trypanothione synthase. |
| 6.1.1.21 Histidine--tRNA ligase. | 6.3.1.10 Adenosylcobinamide-phosphate synthase. |

| | |
|---|---|
| 6.1.1.22 Asparagine--tRNA ligase. | 6.3.2.1 Pantoate--beta-alanine ligase. |
| 6.1.1.23 Aspartate--tRNA(Asn) ligase. | 6.3.2.2 Glutamate--cysteine ligase. |
| 6.1.1.24 Glutamate--tRNA(Gln) ligase. | 6.3.2.3 Glutathione synthase. |
| 6.1.1.25 Lysine--tRNA(Pyl) ligase. | 6.3.2.4 D-alanine--D-alanine ligase. |
| 6.2.1.1 Acetate--CoA ligase. | 6.3.2.5 Phosphopantothenate--cysteine ligase. |
| 6.2.1.2 Butyrate--CoA ligase. | 6.3.2.6 Phosphoribosylaminoimidazolesuccinocarboxamide synthase. |
| 6.2.1.3 Long-chain-fatty-acid--CoA ligase. | 6.3.2.7 UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--L-lysine ligase. |
| 6.2.1.4 Succinate--CoA ligase (GDP-forming). | 6.3.2.8 UDP-N-acetylmuramate--L-alanine ligase. |

| | |
|---|---|
| 6.3.2.9 UDP-N-acetylmuramoylalanine--D-glutamate ligase. | 6.3.4.14 Biotin carboxylase. |
| 6.3.2.10 UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase. | 6.3.4.15 Biotin--[acetyl-CoA-carboxylase] ligase. |
| 6.3.2.11 Carnosine synthase. | 6.3.4.16 Carbamoyl-phosphate synthase (ammonia). |
| 6.3.2.12 Dihydrofolate synthase. | 6.3.4.17 Formate--dihydrofolate ligase. |
| 6.3.2.13 UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase. | 6.3.5.1 NAD(+) synthase (glutamine-hydrolyzing). |
| 6.3.2.14 2,3-dihydroxybenzoate--serine ligase. | 6.3.5.2 GMP synthase (glutamine-hydrolyzing). |
| 6.3.2.16 D-alanine--alanyl-poly(glycerolphosphate) ligase. | 6.3.5.3 Phosphoribosylformylglycinamidine synthase. |
| 6.3.2.17 Tetrahydrofolylpolyglutamate synthase. | 6.3.5.4 Asparagine synthase (glutamine-hydrolyzing). |
| 6.3.2.18 Gamma-glutamylhistamine synthase. | 6.3.5.5 Carbamoyl-phosphate synthase (glutamine-hydrolyzing). |
| 6.3.2.19 Ubiquitin--protein ligase. | 6.3.5.6 Asparaginyl-tRNA synthase (glutamine-hydrolyzing). |
| 6.3.2.20 Indoleacetate-lysine synthetase. | 6.3.5.7 Glutaminyl-tRNA synthase (glutamine-hydrolyzing). |
| 6.3.2.21 Ubiquitin--calmodulin ligase. | 6.3.5.8 Aminodeoxychorismate synthase. |
| 6.3.2.22 Diphthine--ammonia ligase. | 6.3.5.9 Hydrogenobyrinic acid a,c-diamide synthase (glutamine-hydrolyzing). |
| 6.3.2.23 Homoglutathione synthase. | 6.3.5.10 Adenosylcobyric acid synthase (glutamine-hydrolyzing). |
| 6.3.2.24 Tyrosine--arginine ligase. | 6.4.1.1 Pyruvate carboxylase. |
| 6.3.2.25 Tubulin--tyrosine ligase. | 6.4.1.2 Acetyl-CoA carboxylase. |
| 6.3.2.26 N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase. | 6.4.1.3 Propionyl-CoA carboxylase. |
| 6.3.2.27 Aerobactin synthase. | 6.4.1.4 Methylcrotonoyl-CoA carboxylase. |
| 6.3.3.1 Phosphoribosylformylglycinamidine cyclo-ligase. | 6.4.1.5 Geranoyl-CoA carboxylase. |
| 6.3.3.2 5-formyltetrahydrofolate cyclo-ligase. | 6.4.1.6 Acetone carboxylase. |
| 6.3.3.3 Dethiobiotin synthase. | 6.5.1.1 DNA ligase (ATP). |
| 6.3.3.4 (Carboxyethyl)arginine beta-lactam-synthase. | 6.5.1.2 DNA ligase (NAD+). |
| 6.3.4.1 GMP synthase. | 6.5.1.3 RNA ligase (ATP). |
| 6.3.4.2 CTP synthase. | 6.5.1.4 RNA-3'-phosphate cyclase. |
| 6.3.4.3 Formate--tetrahydrofolate ligase. | 6.6.1.1 Magnesium chelatase. |

| | |
|---|---|
| 6.3.4.4 Adenylosuccinate synthase. | 6.6.1.2 Cobaltochelatase. |
| 6.3.4.5 Argininosuccinate synthase. | 6.3.4.17 Formate--dihydrofolate ligase. |
| 6.3.4.6 Urea carboxylase. | 6.3.5.1 NAD(+) synthase (glutamine-hydrolyzing). |
| 6.3.4.7 Ribose-5-phosphate--ammonia ligase. | 6.3.5.2 GMP synthase (glutamine-hydrolyzing). |
| 6.3.4.8 Imidazoleacetate--phosphoribosyldiphosphate ligase. | 6.3.5.3 Phosphoribosylformylglycinamidine synthase. |
| 6.3.4.9 Biotin--[methylmalonyl-CoA-carboxytransferase] ligase. | 6.3.5.4 Asparagine synthase (glutamine-hydrolyzing). |
| 6.3.4.10 Biotin--[propionyl-CoA-carboxylase (ATP-hydrolyzing)] ligase. | 6.3.5.5 Carbamoyl-phosphate synthase (glutamine-hydrolyzing). |
| 6.3.4.11 Biotin--[methylcrotonoyl-CoA-carboxylase] ligase. | 6.3.5.6 Asparaginyl-tRNA synthase (glutamine-hydrolyzing). |
| 6.3.4.12 Glutamate--methylamine ligase. | 6.3.5.7 Glutaminyl-tRNA synthase (glutamine-hydrolyzing). |
| 6.3.4.13 Phosphoribosylamine- glycine ligase. | 6.3.5.8 Aminodeoxychorismate synthase. |

| | |
|---|---|
| 6.3.5.9 Hydrogenobyrinic acid a,c-diamide synthase (glutamine-hydrolyzing). | 6.4.1.6 Acetone carboxylase. |
| 6.3.5.10 Adenosylcobyric acid synthase (glutamine-hydrolyzing). | 6.5.1.1 DNA ligase (ATP). |
| 6.4.1.1 Pyruvate carboxylase. | 6.5.1.2 DNA ligase (NAD+). |
| 6.4.1.2 Acetyl-CoA carboxylase. | 6.5.1.3 RNA ligase (ATP). |
| 6.4.1.3 Propionyl-CoA carboxylase. | 6.5.1.4 RNA-3'-phosphate cyclase. |
| 6.4.1.4 Methylcrotonoyl-CoA carboxylase. | 6.6.1.1 Magnesium chelatase. |
| 6.4.1.5 Geranoyl-CoA carboxylase. | 6.6.1.2 Cobaltochelatase. |

Table 3 summarizes exemplary functions of exemplary polypeptide and/or peptides (e.g., enzymes and active fragments thereof) of the invention; these enzyme functions were determined using sequence identity comparison analysis using closest BLAST hits to the exemplary polypeptides and polynucleotides of the invention.

The invention also provides isolated, synthetic and recombinant nucleic acids encoding polypeptides, e.g., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, etc., and all additional nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699 (the exemplary polynucleotides of the invention). The invention also provides isolated, synthetic and recombinant polypeptides, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699 (the exemplary polypeptides of the invention).

In another embodiment, the polypeptides of the invention can be expressed in any expression system, in vitro or in vivo, e.g., any microorganism or other cell system (e.g.,

eukaryotic, such as yeast or mammalian cells) using procedures known in the art. In other aspects, the polypeptides of the invention can be immobilized on a solid support prior to use in the methods of the invention. Methods for immobilizing enzymes on solid supports are commonly known in the art, for example J. Mol. Cat. B: Enzymatic 6 (1999) 29-39; Chivata et al. Biocatalysis: Immobilized cells and enzymes, J Mol. Cat. 37 (1986) 1-24: Sharma et al., Immobilized Biomaterials Techniques and Applications, Angew. Chem. Int. Ed. Engl. 21 (1982) 837-54: Laskin (Ed.), Enzymes and Immobilized Cells in Biotechnology.

Nucleic Acids

The invention provides nucleic acids (e.g., the various genuses of polynucleotides based in the exemplary SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, etc., including all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699), including expression cassettes such as expression vectors, encoding polypeptides (e.g., enzymes) of the invention. The invention also includes methods for discovering new polypeptide (e.g., enzyme) sequences using the nucleic acids of the invention. The invention also includes methods for inhibiting the expression of enzymes, genes, transcripts and polypeptides using the nucleic acids of the invention. Also provided are methods for modifying the nucleic acids of the invention by, e.g., synthetic ligation reassembly, optimized directed evolution system and/or saturation mutagenesis.

The nucleic acids of the invention can be made, isolated and/or manipulated by, e.g., cloning and expression of cDNA libraries, amplification of message or genomic DNA by PCR, and the like. For example, exemplary sequences of the invention were initially derived from environmental sources.

In one aspect, the invention provides nucleic acids, and the polypeptides encoded by them, with a common novelty in that they are derived from a common source, e.g., an environmental or a bacterial source.

In practicing the methods of the invention, homologous genes can be modified by manipulating a template nucleic acid, as described herein. The invention can be practiced in conjunction with any method or protocol or device known in the art, which are well described in the scientific and patent literature.

The phrases "nucleic acid" or "nucleic acid sequence" as used herein refer to an oligonucleotide, nucleotide, polynucleotide, or to a fragment of any of these, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded

and may represent a sense or antisense (complementary) strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material, natural or synthetic in origin. The phrases "nucleic acid" or "nucleic acid sequence" includes oligonucleotide, nucleotide, polynucleotide, or to a fragment of any of these, to DNA or RNA (e.g., mRNA, rRNA, tRNA, iRNA) of genomic or synthetic origin which may be single-stranded or double-stranded and may represent a sense or antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material, natural or synthetic in origin, including, e.g., iRNA, ribonucleoproteins (e.g., e.g., double stranded iRNAs, e.g., iRNPs). The term encompasses nucleic acids, i.e., oligonucleotides, containing known analogues of natural nucleotides. The term also encompasses nucleic-acid-like structures with synthetic backbones, see e.g., Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197; Strauss-Soukup (1997) Biochemistry 36:8692-8698; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156. "Oligonucleotide" includes either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide can ligate to a fragment that has not been dephosphorylated.

In one aspect, "coding sequence of" or a "nucleotide sequence encoding" a particular polypeptide or protein, is a nucleic acid sequence which is transcribed and translated into a polypeptide or protein when placed under the control of appropriate regulatory sequences. The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as, where applicable, intervening sequences (introns) between individual coding segments (exons). "Operably linked" as used herein refers to a functional relationship between two or more nucleic acid (e.g., DNA) segments. Typically, it refers to the functional relationship of transcriptional regulatory sequence to a transcribed sequence. For example, a promoter is operably linked to a coding sequence, such as a nucleic acid of the invention, if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be

physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

The phrase "substantially identical" in the context of two nucleic acids or polypeptides, refers to two or more sequences that have, e.g., at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more nucleotide or amino acid residue (sequence) identity, when compared and aligned for maximum correspondence, as measured using one of the known sequence comparison algorithms or by visual inspection. In alternative aspects, the substantial identity exists over a region of at least about 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, or more residues or alternatively the sequences are substantially identical over at least about 150 to 200 or more residues, or over the full length of the gene and/or the protein coding sequence. In some aspects, the sequences are substantially identical over the entire length of the coding regions.

The term "variant" refers to polynucleotides or polypeptides of the invention modified at one or more base pairs, codons, introns, exons, or amino acid residues (respectively) yet still retain the biological activity of a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention. Variants can be produced by any number of means included methods such as, for example, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, GSSM and any combination thereof.

The term "saturation mutagenesis", Gene Site Saturation Mutagenesis, or "GSSM" includes a method that uses degenerate oligonucleotide primers to introduce point mutations into a polynucleotide, as described in detail, below.

The term "optimized directed evolution system" or "optimized directed evolution" includes a method for reassembling fragments of related nucleic acid sequences, e.g., related genes, and explained in detail, below.

The term "synthetic ligation reassembly" or "SLR" includes a method of ligating oligonucleotide fragments in a non-stochastic fashion, and explained in detail, below.

As used herein, the term "promoter" includes all sequences capable of driving transcription of a coding sequence in a cell, e.g., a plant cell. Thus, promoters used in the constructs of the invention include *cis*-acting transcriptional control elements and regulatory sequences that are involved in regulating or modulating the timing and/or rate of transcription of a gene. For example, a promoter can be a *cis*-acting transcriptional control element, including an enhancer, a promoter, a transcription terminator, an origin of replication, a chromosomal integration sequence, 5' and 3' untranslated regions, or an intronic sequence, which are involved in transcriptional regulation. These cis-acting sequences typically interact with proteins or other biomolecules to carry out (turn on/off, regulate, modulate, etc.) transcription. "Constitutive" promoters are those that drive expression continuously under most environmental conditions and states of development or cell differentiation. "Inducible" or "regulatable" promoters direct expression of the nucleic acid of the invention under the influence of environmental conditions or developmental conditions. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, drought, or the presence of light.

"Plasmids" can be commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. Equivalent plasmids to those described herein are known in the art and will be apparent to the ordinarily skilled artisan.

In one aspect, the term "recombinant" means that the nucleic acid is adjacent to a "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the nucleic acids will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Typically, the enriched nucleic acids represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More typically, the enriched nucleic acids represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a one aspect, the enriched nucleic acids represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

One aspect of the invention is an isolated nucleic acid comprising one of the sequences of the invention, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40,

50, 75, 100, 150, 200, 300, 400, or 500 or more consecutive bases of a nucleic acid of the invention. The isolated, nucleic acids may comprise DNA, including cDNA, genomic DNA and synthetic DNA. The DNA may be double-stranded or single-stranded and if single stranded may be the coding strand or non-coding (anti-sense) strand. Alternatively, the isolated nucleic acids may comprise RNA.

The isolated nucleic acids of the invention may be used to prepare one of the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids of one of the polypeptides of the invention. Accordingly, another aspect of the invention is an isolated nucleic acid which encodes one of the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids of one of the polypeptides of the invention. The coding sequences of these nucleic acids may be identical to one of the coding sequences of one of the nucleic acids of the invention or may be different coding sequences which encode one of the of the invention having at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids of one of the polypeptides of the invention, as a result of the redundancy or degeneracy of the genetic code. The genetic code is well known to those of skill in the art and can be obtained, e.g., on page 214 of B. Lewin, Genes VI, Oxford University Press, 1997.

The isolated nucleic acid which encodes one of the polypeptides of the invention, but is not limited to: only the coding sequence of a nucleic acid of the invention and additional coding sequences, such as leader sequences or proprotein sequences and non-coding sequences, such as introns or non-coding sequences 5' and/or 3' of the coding sequence. Thus, as used herein, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only the coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

Alternatively, the nucleic acid sequences of the invention, may be mutagenized using conventional techniques, such as site directed mutagenesis, or other techniques familiar to those skilled in the art, to introduce silent changes into the polynucleotides o of the invention. As used herein, "silent changes" include, for example, changes which do not alter the amino acid sequence encoded by the polynucleotide. Such changes may be desirable in order to increase the level of the polypeptide produced by host cells containing a vector encoding the polypeptide by introducing codons or codon pairs which occur frequently in the host organism.

The invention also relates to polynucleotides which have nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptides of the invention. Such nucleotide changes may be introduced using techniques such as site directed mutagenesis, random chemical mutagenesis, exonuclease III deletion and other recombinant DNA techniques. Alternatively, such nucleotide changes may be naturally occurring allelic variants which are isolated by identifying nucleic acids which specifically hybridize to probes comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of the invention (or the sequences complementary thereto) under conditions of high, moderate, or low stringency as provided herein.

*General Techniques*

The nucleic acids used to practice this invention, whether RNA, iRNA, antisense nucleic acid, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed/ generated recombinantly. Recombinant polypeptides generated from these nucleic acids can be individually isolated, synthetic or cloned and tested for a desired activity. Any recombinant expression system can be used, including bacterial, mammalian, yeast, insect or plant cell expression systems.

Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Patent No. 4,458,066.

Techniques for the manipulation of nucleic acids, such as, e.g., subcloning, labeling probes (e.g., random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, hybridization and the like are well described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Another useful means of obtaining and manipulating nucleic acids used to practice the methods of the invention is to clone from genomic samples, and, if desired, screen and re-clone inserts isolated, synthetic or amplified from, e.g., genomic clones or cDNA clones. Sources of nucleic acid used in the methods of the invention include genomic or cDNA libraries contained in, e.g., mammalian artificial chromosomes (MACs), see, e.g., U.S. Patent Nos. 5,721,118; 6,025,155; human artificial chromosomes, see, e.g., Rosenfeld (1997) Nat. Genet. 15:333-335; yeast artificial chromosomes (YAC); bacterial artificial chromosomes (BAC); P1 artificial chromosomes, see, e.g., Woon (1998) Genomics 50:306-316; P1-derived vectors (PACs), see, e.g., Kern (1997) Biotechniques 23:120-124; cosmids, recombinant viruses, phages or plasmids.

In one aspect, a nucleic acid encoding a polypeptide of the invention is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptide or fragment thereof.

The invention provides fusion proteins and nucleic acids encoding them. A polypeptide of the invention can be fused to a heterologous peptide or polypeptide, such as N-terminal identification peptides which impart desired characteristics, such as increased stability or simplified purification. Peptides and polypeptides of the invention can also be synthesized and expressed as fusion proteins with one or more additional domains linked thereto for, e.g., producing a more immunogenic peptide, to more readily isolate a recombinantly synthesized peptide, to identify and isolate antibodies and antibody-expressing B cells, and the like. Detection and purification facilitating domains include, e.g., metal chelating peptides such as polyhistidine tracts and histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego CA) between a purification domain and the motif-comprising peptide or polypeptide to facilitate purification. For example, an expression vector can include an epitope-encoding nucleic acid sequence linked to six histidine residues followed by a thioredoxin and an enterokinase cleavage site (see e.g., Williams (1995) Biochemistry 34:1787-1797; Dobeli (1998) Protein Expr. Purif. 12:404-414). The histidine residues facilitate detection and purification while the enterokinase cleavage site provides a means for purifying the epitope from the remainder of the fusion protein. Technology pertaining

to vectors encoding fusion proteins and application of fusion proteins are well described in the scientific and patent literature, see e.g., Kroll (1993) DNA Cell. Biol., 12:441-53.

*Transcriptional and translational control sequences*

The invention provides nucleic acid (e.g., DNA) sequences of the invention operatively linked to expression (e.g., transcriptional or translational) control sequence(s), e.g., promoters or enhancers, to direct or modulate RNA synthesis/ expression. The expression control sequence can be in an expression vector. Exemplary bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Exemplary eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein I.

Promoters suitable for expressing a polypeptide in bacteria include the *E. coli* lac or trp promoters, the lacI promoter, the lacZ promoter, the T3 promoter, the T7 promoter, the gpt promoter, the lambda PR promoter, the lambda PL promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses, and the mouse metallothionein-I promoter. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used. Promoters suitable for expressing the polypeptide or fragment thereof in bacteria include the *E. coli lac* or *trp* promoters, the *lacI* promoter, the *lacZ* promoter, the *T3* promoter, the *T7* promoter, the *gpt* promoter, the *lambda* $P_R$ promoter, the *lambda* $P_L$ promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK) and the acid phosphatase promoter. Fungal promoters include the α-factor promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses and the mouse metallothionein-I promoter. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used.

*Tissue-Specific Promoters*

The invention provides expression cassettes that can be expressed in a tissue-specific manner, e.g., that can express a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention in a tissue-specific manner. The invention also provides plants or seeds that express a polypeptide, enzyme, protein, e.g. structural or binding

protein, of the invention in a tissue-specific manner. The tissue-specificity can be seed specific, stem specific, leaf specific, root specific, fruit specific and the like.

The term "expression cassette" as used herein refers to a nucleotide sequence which is capable of affecting expression of a structural gene (i.e., a protein coding sequence, such as a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention) in a host compatible with such sequences. Expression cassettes include at least a promoter operably linked with the polypeptide coding sequence; and, optionally, with other sequences, e.g., transcription termination signals. Additional factors necessary or helpful in effecting expression may also be used, e.g., enhancers, alpha-factors. Thus, expression cassettes also include plasmids, expression vectors, recombinant viruses, any form of recombinant "naked DNA" vector, and the like. A "vector" comprises a nucleic acid which can infect, transfect, transiently or permanently transduce a cell. It will be recognized that a vector can be a naked nucleic acid, or a nucleic acid complexed with protein or lipid. The vector optionally comprises viral or bacterial nucleic acids and/or proteins, and/or membranes (e.g., a cell membrane, a viral lipid envelope, etc.). Vectors include, but are not limited to replicons (e.g., RNA replicons, bacteriophages) to which fragments of DNA may be attached and become replicated. Vectors thus include, but are not limited to RNA, autonomous self-replicating circular or linear DNA or RNA (e.g., plasmids, viruses, and the like, see, e.g., U.S. Patent No. 5,217,879), and include both the expression and non-expression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression vector" this includes both extra-chromosomal circular and linear DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or is incorporated within the host's genome.

"Tissue-specific" promoters are transcriptional control elements that are only active in particular cells or tissues or organs, e.g., in plants or animals. Tissue-specific regulation may be achieved by certain intrinsic factors which ensure that genes encoding proteins specific to a given tissue are expressed. Such factors are known to exist in mammals and plants so as to allow for specific tissues to develop.

The term "plant" includes whole plants, plant parts (e.g., leaves, stems, flowers, roots, etc.), plant protoplasts, seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms

(monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous states. As used herein, the term "transgenic plant" includes plants or plant cells into which a heterologous nucleic acid sequence has been inserted, e.g., the nucleic acids and various recombinant constructs (e.g., expression cassettes) of the invention.

In one aspect, a constitutive promoter such as the CaMV 35S promoter can be used for expression in specific parts of the plant or seed or throughout the plant. For example, for overexpression, a plant promoter fragment can be employed which will direct expression of a nucleic acid in some or all tissues of a plant, e.g., a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such genes include, e.g., *ACT11* from *Arabidopsis* (Huang (1996) *Plant Mol. Biol.* 33:125-139); *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong (1996) *Mol. Gen. Genet.* 251:196-203); the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe (1994) *Plant Physiol.* 104:1167-1176); *GPc1* from maize (GenBank No. X15596; Martinez (1989) *J. Mol. Biol* 208:551-565); the *Gpc2* from maize (GenBank No. U45855, Manjunath (1997) *Plant Mol. Biol.* 33:97-112); plant promoters described in U.S. Patent Nos. 4,962,028; 5,633,440.

The invention uses tissue-specific or constitutive promoters derived from viruses which can include, *e.g.*, the tobamovirus subgenomic promoter (Kumagai (1995) Proc. Natl. Acad. Sci. USA 92:1679-1683; the rice tungro bacilliform virus (RTBV), which replicates only in phloem cells in infected rice plants, with its promoter which drives strong phloem-specific reporter gene expression; the cassava vein mosaic virus (CVMV) promoter, with highest activity in vascular elements, in leaf mesophyll cells, and in root tips (Verdaguer (1996) Plant Mol. Biol. 31:1129-1139).

Alternatively, the plant promoter may direct expression of a polypeptide, enzyme, protein, e.g. structural or binding protein-expressing nucleic acid in a specific tissue, organ or cell type (*i.e.* tissue-specific promoters) or may be otherwise under more precise environmental or developmental control or under the control of an inducible promoter. Examples of environmental conditions that may affect transcription include anaerobic

conditions, elevated temperature, the presence of light, or sprayed with chemicals/hormones. For example, the invention incorporates the drought-inducible promoter of maize (Busk (1997) supra); the cold, drought, and high salt inducible promoter from potato (Kirch (1997) Plant Mol. Biol. 33:897-909).

Tissue-specific promoters can promote transcription only within a certain time frame of developmental stage within that tissue. See, e.g., Blazquez (1998) Plant Cell 10:791-800, characterizing the *Arabidopsis* LEAFY gene promoter. See also Cardon (1997) *Plant J* 12:367-77, describing the transcription factor SPL3, which recognizes a conserved sequence motif in the promoter region of the *A. thaliana* floral meristem identity gene AP1; and Mandel (1995) Plant Molecular Biology, Vol. 29, pp 995-1004, describing the meristem promoter eIF4. Tissue specific promoters which are active throughout the life cycle of a particular tissue can be used. In one aspect, the nucleic acids of the invention are operably linked to a promoter active primarily only in cotton fiber cells. In one aspect, the nucleic acids of the invention are operably linked to a promoter active primarily during the stages of cotton fiber cell elongation, e.g., as described by Rinehart (1996) supra. The nucleic acids can be operably linked to the Fbl2A gene promoter to be preferentially expressed in cotton fiber cells (Ibid) . See also, John (1997) Proc. Natl. Acad. Sci. USA 89:5769-5773; John, et al., U.S. Patent Nos. 5,608,148 and 5,602,321, describing cotton fiber-specific promoters and methods for the construction of transgenic cotton plants. Root-specific promoters may also be used to express the nucleic acids of the invention. Examples of root-specific promoters include the promoter from the alcohol dehydrogenase gene (DeLisle (1990) Int. Rev. Cytol. 123:39-60). Other promoters that can be used to express the nucleic acids of the invention include, e.g., ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed coat-specific promoters, or some combination thereof; a leaf-specific promoter (see, e.g., Busk (1997) Plant J. 11:1285 1295, describing a leaf-specific promoter in maize); the ORF13 promoter from *Agrobacterium rhizogenes* (which exhibits high activity in roots, see, e.g., Hansen (1997) supra); a maize pollen specific promoter (see, e.g., Guerrero (1990) Mol. Gen. Genet. 224:161 168); a tomato promoter active during fruit ripening, senescence and abscission of leaves and, to a lesser extent, of flowers can be used (see, e.g., Blume (1997) Plant J. 12:731 746); a pistil-specific promoter from the potato SK2 gene (see, e.g., Ficker (1997) Plant Mol. Biol. 35:425 431); the Blec4 gene from pea, which is active in epidermal tissue of vegetative and floral shoot apices of transgenic alfalfa making it a useful tool to target the expression of

foreign genes to the epidermal layer of actively growing shoots or fibers; the ovule-specific BEL1 gene (see, e.g., Reiser (1995) Cell 83:735-742, GenBank No. U39944); and/or, the promoter in Klee, U.S. Patent No. 5,589,583, describing a plant promoter region is capable of conferring high levels of transcription in meristematic tissue and/or rapidly dividing cells.

Alternatively, plant promoters which are inducible upon exposure to plant hormones, such as auxins, are used to express the nucleic acids of the invention. For example, the invention can use the auxin-response elements E1 promoter fragment (AuxREs) in the soybean (*Glycine max* L.) (Liu (1997) Plant Physiol. 115:397-407); the auxin-responsive *Arabidopsis* GST6 promoter (also responsive to salicylic acid and hydrogen peroxide) (Chen (1996) Plant J. 10: 955-966); the auxin-inducible parC promoter from tobacco (Sakai (1996) 37:906-913); a plant biotin response element (Streit (1997) Mol. Plant Microbe Interact. 10:933-937); and, the promoter responsive to the stress hormone abscisic acid (Sheen (1996) Science 274:1900-1902).

The nucleic acids of the invention can also be operably linked to plant promoters which are inducible upon exposure to chemicals reagents which can be applied to the plant, such as herbicides or antibiotics. For example, the maize In2-2 promoter, activated by benzenesulfonamide herbicide safeners, can be used (De Veylder (1997) Plant Cell Physiol. 38:568-577); application of different herbicide safeners induces distinct gene expression patterns, including expression in the root, hydathodes, and the shoot apical meristem. Coding sequence can be under the control of, *e.g.*, a tetracycline-inducible promoter, *e.g.*, as described with transgenic tobacco plants containing the *Avena sativa* L. (oat) arginine decarboxylase gene (Masgrau (1997) Plant J. 11:465-473); or, a salicylic acid-responsive element (Stange (1997) Plant J. 11:1315-1324). Using chemically- (*e.g.*, hormone- or pesticide-) induced promoters, *i.e.*, promoter responsive to a chemical which can be applied to the transgenic plant in the field, expression of a polypeptide of the invention can be induced at a particular stage of development of the plant. Thus, the invention also provides for transgenic plants containing an inducible gene encoding for polypeptides of the invention whose host range is limited to target plant species, such as corn, rice, barley, wheat, potato or other crops, inducible at any stage of development of the crop.

One of skill will recognize that a tissue-specific plant promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, a

tissue-specific promoter is one that drives expression preferentially in the target tissue or cell type, but may also lead to some expression in other tissues as well.

The nucleic acids of the invention can also be operably linked to plant promoters which are inducible upon exposure to chemicals reagents. These reagents include, e.g., herbicides, synthetic auxins, or antibiotics which can be applied, e.g., sprayed, onto transgenic plants. Inducible expression of the polypeptide, enzyme, protein, e.g. structural or binding protein-producing nucleic acids of the invention will allow the grower to select plants with the optimal polypeptide, enzyme, protein, e.g. structural or binding protein, expression and/or activity. The development of plant parts can thus controlled. In this way the invention provides the means to facilitate the harvesting of plants and plant parts. For example, in various embodiments, the maize In2-2 promoter, activated by benzenesulfonamide herbicide safeners, is used (De Veylder (1997) Plant Cell Physiol. 38:568-577); application of different herbicide safeners induces distinct gene expression patterns, including expression in the root, hydathodes, and the shoot apical meristem. Coding sequences of the invention are also under the control of a tetracycline-inducible promoter, e.g., as described with transgenic tobacco plants containing the *Avena sativa* L. (oat) arginine decarboxylase gene (Masgrau (1997) Plant J. 11:465-473); or, a salicylic acid-responsive element (Stange (1997) Plant J. 11:1315-1324).

In some aspects, proper polypeptide expression may require polyadenylation region at the 3'-end of the coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant (or animal or other) genes, or from genes in the *Agrobacterial* T-DNA.

*Expression vectors and cloning vehicles*

The invention provides expression vectors and cloning vehicles comprising nucleic acids of the invention, e.g., sequences encoding the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention. Expression vectors and cloning vehicles of the invention can comprise viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral DNA (e.g., vaccinia, adenovirus, foul pox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus, Aspergillus and yeast). Vectors of the invention can include chromosomal, non-chromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and

are commercially available. Exemplary vectors are include: bacterial: pQE vectors (Qiagen), pBLUESCRIPT plasmids, pNH vectors, (lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as they are replicable and viable in the host. Low copy number or high copy number vectors may be employed with the present invention.

The expression vector can comprise a promoter, a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Mammalian expression vectors can comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences. In some aspects, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required non-transcribed genetic elements.

In one aspect, the expression vectors contain one or more selectable marker genes to permit selection of host cells containing the vector. Such selectable markers include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli*, and the *S. cerevisiae* TRP1 gene. Promoter regions can be selected from any desired gene using chloramphenicol transferase (CAT) vectors or other vectors with selectable markers.

Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells can also contain enhancers to increase expression levels. Enhancers are cis-acting elements of DNA that can be from about 10 to about 300 bp in length. They can act on a promoter to increase its transcription. Exemplary enhancers include the SV40 enhancer on the late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and the adenovirus enhancers.

A nucleic acid sequence can be inserted into a vector by a variety of procedures. In general, the sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, blunt ends in both the insert and the vector may be ligated. A variety of cloning techniques are known in the art, e.g., as described in Ausubel and

Sambrook. Such procedures and others are deemed to be within the scope of those skilled in the art.

The vector can be in the form of a plasmid, a viral particle, or a phage. Other vectors include chromosomal, non-chromosomal and synthetic DNA sequences, derivatives of SV40; bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. A variety of cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by, e.g., Sambrook.

Particular bacterial vectors which can be used include the commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), GEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174 pBLUESCRIPT II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, DR540, pRIT5 (Pharmacia), pKK232-8 and pCM7. Particular eukaryotic vectors include pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). However, any other vector may be used as long as it is replicable and viable in the host cell.

The nucleic acids of the invention can be expressed in expression cassettes, vectors or viruses and transiently or stably expressed in plant cells and seeds. One exemplary transient expression system uses episomal expression systems, e.g., cauliflower mosaic virus (CaMV) viral RNA generated in the nucleus by transcription of an episomal mini-chromosome containing supercoiled DNA, see, e.g., Covey (1990) Proc. Natl. Acad. Sci. USA 87:1633-1637. Alternatively, coding sequences, i.e., all or sub-fragments of sequences of the invention can be inserted into a plant host cell genome becoming an integral part of the host chromosomal DNA. Sense or antisense transcripts can be expressed in this manner. A vector comprising the sequences (e.g., promoters or coding regions) from nucleic acids of the invention can comprise a marker gene that confers a selectable phenotype on a plant cell or a seed. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

Expression vectors capable of expressing nucleic acids and proteins in plants are well known in the art, and can include, *e.g.*, vectors from *Agrobacterium* spp., potato virus X (see, *e.g.*, Angell (1997) EMBO J. 16:3675-3684), tobacco mosaic virus

(see, *e.g.*, Casper (1996) Gene 173:69-73), tomato bushy stunt virus (see, *e.g.*, Hillman (1989) Virology 169:42-50), tobacco etch virus (see, *e.g.*, Dolja (1997) Virology 234:243-252), bean golden mosaic virus (see, *e.g.*, Morinaga (1993) Microbiol Immunol. 37:471-476), cauliflower mosaic virus (see, *e.g.*, Cecchini (1997) Mol. Plant Microbe Interact. 10:1094-1101), maize Ac/Ds transposable element (see, *e.g.*, Rubin (1997) Mol. Cell. Biol. 17:6294-6302; Kunze (1996) Curr. Top. Microbiol. Immunol. 204:161-194), and the maize suppressor-mutator (Spm) transposable element (see, *e.g.*, Schlappi (1996) Plant Mol. Biol. 32:717-725); and derivatives thereof.

In one aspect, the expression vector can have two replication systems to allow it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector can contain at least one sequence homologous to the host cell genome. It can contain two homologous sequences which flank the expression construct. The integrating vector can be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art.

Expression vectors of the invention may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed, e.g., genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers can also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct RNA synthesis. Particular named bacterial promoters include *lacI*, *lacZ*, *T3*, *T7*, *gpt*, *lambda* $P_R$, $P_L$ and *trp*. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Promoter regions can be selected from any desired gene using chloramphenicol transferase (CAT) vectors or other vectors with selectable markers. In addition, the expression vectors in one aspect contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as

dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Mammalian expression vectors may also comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences and 5' flanking nontranscribed sequences. In some aspects, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells may also contain enhancers to increase expression levels. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp in length that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin and the adenovirus enhancers.

In addition, the expression vectors typically contain one or more selectable marker genes to permit selection of host cells containing the vector. Such selectable markers include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli* and the *S. cerevisiae TRP1* gene.

In some aspects, the nucleic acid encoding one of the polypeptides of the invention, or fragments comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptide or fragment thereof. Optionally, the nucleic acid can encode a fusion polypeptide in which one of the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is fused to heterologous peptides or polypeptides, such as N-terminal identification peptides which impart desired characteristics, such as increased stability or simplified purification.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, blunt ends in both the insert and the vector may be ligated. A variety of cloning techniques are disclosed in Ausubel *et al.* Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook *et al.*, Molecular

Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory Press (1989. Such procedures and others are deemed to be within the scope of those skilled in the art.

The vector may be, for example, in the form of a plasmid, a viral particle, or a phage. Other vectors include chromosomal, nonchromosomal and synthetic DNA sequences, derivatives of SV40; bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus and pseudorabies. A variety of cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor, N.Y., (1989).

*Host cells and transformed cells*

The invention also provides a transformed cell comprising a nucleic acid sequence of the invention, e.g., a sequence encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention, or a vector of the invention. The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells, eukaryotic cells, such as bacterial cells, fungal cells, yeast cells, mammalian cells, insect cells, or plant cells. Exemplary bacterial cells include any species of *Streptomyces*, *Staphylococcus*, *Pseudomonas* or *Bacillus*, including *E. coli*, *Bacillus subtilis*, *Pseudomonas fluorescens*, *Bacillus cereus,* or *Salmonella typhimurium*. Exemplary fungal cells include any species of *Aspergillus*. Exemplary yeast cells include any species of *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Schwanniomyces*, including *Pichia pastoris*, *Saccharomyces cerevisiae*, or *Schizosaccharomyces pombe*. Exemplary insect cells include any species of *Spodoptera* or *Drosophila*, including *Drosophila S2* and *Spodoptera Sf9*. Exemplary animal cells include CHO, COS or Bowes melanoma or any mouse or human cell line. The selection of an appropriate host is within the abilities of those skilled in the art. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g., Weising (1988) Ann. Rev. Genet. 22:421-477; U.S. Patent No. 5,750,870.

The vector can be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

In one aspect, the nucleic acids or vectors of the invention are introduced into the cells for screening, thus, the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type. Exemplary methods include $CaPO_4$ precipitation, liposome fusion, lipofection (e.g., LIPOFECTIN™), electroporation, viral infection, etc. The candidate nucleic acids may stably integrate into the genome of the host cell (for example, with retroviral introduction) or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.). As many pharmaceutically important screens require human or model mammalian cell targets, retroviral vectors capable of transfecting such targets can be used.

Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (e.g., temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or fragment thereof.

Cells can be harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Depending upon the host employed in a recombinant production procedure, the polypeptides produced by host cells

containing the vector may be glycosylated or may be non-glycosylated. Polypeptides of the invention may or may not also include an initial methionine amino acid residue.

Cell-free translation systems can also be employed to produce a polypeptide of the invention. Cell-free translation systems can use mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment thereof. In some aspects, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

The expression vectors can contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Host cells containing the polynucleotides of interest, e.g., nucleic acids of the invention, can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression and will be apparent to the ordinarily skilled artisan. The clones which are identified as having the specified enzyme activity, binding activity and/or structural activity may then be sequenced to identify the polynucleotide sequence encoding an enzyme having the enhanced enzyme activity, binding activity and/or structural activity.

The invention provides a method for overexpressing a recombinant polypeptide, enzyme, protein, e.g. structural or binding protein, in a cell comprising expressing a vector comprising a nucleic acid of the invention, e.g., a nucleic acid comprising a nucleic acid sequence with at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to an exemplary sequence of the invention over a region of at least about 100 residues, wherein the sequence identities are determined by analysis with a sequence comparison algorithm or by visual inspection, or, a nucleic acid that hybridizes under stringent conditions to a nucleic acid sequence of the invention. The overexpression can be

effected by any means, e.g., use of a high activity promoter, a dicistronic vector or by gene amplification of the vector.

The nucleic acids of the invention can be expressed, or overexpressed, in any in vitro or in vivo expression system. Any cell culture systems can be employed to express, or over-express, recombinant protein, including bacterial, insect, yeast, fungal or mammalian cultures. Over-expression can be effected by appropriate choice of promoters, enhancers, vectors (e.g., use of replicon vectors, dicistronic vectors (see, e.g., Gurtu (1996) Biochem. Biophys. Res. Commun. 229:295-8), media, culture systems and the like. In one aspect, gene amplification using selection markers, e.g., glutamine synthetase (see, e.g., Sanders (1987) Dev. Biol. Stand. 66:55-63), in cell systems are used to overexpress the polypeptides of the invention.

The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells, eukaryotic cells, mammalian cells, insect cells, or plant cells. As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces* and *Staphylococcus*, fungal cells, such as *Aspergillus*, yeast such as any species of *Pichia, Saccharomyces, Schizosaccharomyces, Schwanniomyces,* including *Pichia pastoris, Saccharomyces cerevisiae,* or *Schizosaccharomyces pombe*, insect cells such as *Drosophila S2* and *Spodoptera Sf9*, animal cells such as CHO, COS or Bowes melanoma and adenoviruses. The selection of an appropriate host is within the abilities of those skilled in the art.

The vector may be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (*e.g.*, temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or fragment thereof.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts (described by Gluzman, Cell, 23:175, 1981) and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Depending upon the host employed in a recombinant production procedure, the polypeptides produced by host cells containing the vector may be glycosylated or may be non-glycosylated. Polypeptides of the invention may or may not also include an initial methionine amino acid residue.

Alternatively, the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids thereof can be synthetically produced by conventional peptide synthesizers. In other aspects, fragments or portions of the polypeptides may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.

Cell-free translation systems can also be employed to produce one of the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids thereof using mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment thereof. In some aspects, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is

then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

*Amplification of Nucleic Acids*

In practicing the invention, nucleic acids encoding the polypeptides of the invention, or modified nucleic acids, can be reproduced by, e.g., amplification. The invention provides amplification primer sequence pairs for amplifying nucleic acids encoding polypeptides (e.g., enzymes) of the invention. In one aspect, the primer pairs are capable of amplifying nucleic acid sequences of the invention, e.g., including the exemplary SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, etc., including all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, or a subsequence thereof, etc. One of skill in the art can design amplification primer sequence pairs for any part of or the full length of these sequences.

In one aspect, the invention provides a nucleic acid amplified by a primer pair of the invention, e.g., a primer pair as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more residues of a nucleic acid of the invention, and about the first (the 5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more residues of the complementary strand.

The invention provides an amplification primer sequence pair for amplifying a nucleic acid encoding a polypeptide having an enzyme, structural or binding activity, wherein the primer pair is capable of amplifying a nucleic acid comprising a sequence of the invention, or fragments or subsequences thereof. One or each member of the amplification primer sequence pair can comprise an oligonucleotide comprising at least about 10 to 50 or more consecutive bases of the sequence, or about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more consecutive bases of the sequence. The invention provides amplification primer pairs, wherein the primer pair comprises a first member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more residues of a nucleic acid of the invention, and a second member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more residues of the complementary strand of the first member. The invention provides a polypeptide, enzyme, protein, e.g. structural or binding protein, generated by amplification, e.g., polymerase chain reaction (PCR), using an amplification primer pair of the invention. The invention provides methods of making a polypeptide, enzyme, protein, e.g. structural or binding protein, by

amplification, e.g., polymerase chain reaction (PCR), using an amplification primer pair of the invention. In one aspect, the amplification primer pair amplifies a nucleic acid from a library, e.g., a gene library, such as an environmental library.

Amplification reactions can also be used to quantify the amount of nucleic acid in a sample (such as the amount of message in a cell sample), label the nucleic acid (e.g., to apply it to an array or a blot), detect the nucleic acid, or quantify the amount of a specific nucleic acid in a sample. In one aspect of the invention, message isolated from a cell or a cDNA library are amplified.

The skilled artisan can select and design suitable oligonucleotide amplification primers. Amplification methods are also well known in the art, and include, e.g., polymerase chain reaction, PCR (see, e.g., PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR) (see, e.g., Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); transcription amplification (see, e.g., Kwoh (1989) Proc. Natl. Acad. Sci. USA 86:1173); and, self-sustained sequence replication (see, e.g., Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); Q Beta replicase amplification (see, e.g., Smith (1997) J. Clin. Microbiol. 35:1477-1491), automated Q-beta replicase amplification assay (see, e.g., Burg (1996) Mol. Cell. Probes 10:257-271) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario); see also Berger (1987) Methods Enzymol. 152:307-316; Sambrook; Ausubel; U.S. Patent Nos. 4,683,195 and 4,683,202; Sooknanan (1995) Biotechnology 13:563-564.

Determining the degree of sequence identity

The invention provides nucleic acids comprising sequences having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to an exemplary nucleic acid of the invention (e.g., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, etc., including all nucleic acids disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, and nucleic acids encoding SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from

SEQ ID NO:2 through SEQ ID NO:108,699) over a region of at least about 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550 or more, residues. The invention provides polypeptides comprising sequences having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to an exemplary polypeptide of the invention. The extent of sequence identity (homology) may be determined using any computer program and associated parameters, including those described herein, such as BLAST 2.2.2. or FASTA version 3.0t78, with the default parameters.

As used herein, the terms "computer," "computer program" and "processor" are used in their broadest general contexts and incorporate all such devices, as described in detail, below.

Nucleic acid sequences of the invention can comprise at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 or more consecutive nucleotides of an exemplary sequence of the invention and sequences substantially identical thereto. Homologous sequences and fragments of nucleic acid sequences of the invention can refer to a sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity (homology) to these sequences. Homology (sequence identity) may be determined using any of the computer programs and parameters described herein, including FASTA version 3.0t78 with the default parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the nucleic acid sequences of the invention. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error. It will be appreciated that the nucleic acid sequences of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. Biochemistry, 3rd Ed., W. H Freeman & Co., New York.) or in any other format which records the identity of the nucleotides in a sequence.

Various sequence comparison programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention. Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA and CLUSTALW (see, e.g., Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85(8):2444-2448, 1988; Altschul *et al.*, J. Mol. Biol. 215(3):403-410, 1990; Thompson Nucleic Acids Res. 22(2):4673-4680, 1994; Higgins *et al.*, Methods Enzymol. 266:383-402, 1996; Altschul *et al.*, J. Mol. Biol. 215(3):403-410, 1990; Altschul *et al.*, Nature Genetics 3:266-272, 1993).

Homology or identity is often measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various deletions, substitutions and other modifications. The terms "homology" and "identity" in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region as measured using any number of sequence comparison algorithms or by manual alignment and visual inspection.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, Adv. Appl. Math.

2:482, 1981, by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol 48:443, 1970, by the search for similarity method of person & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988, by computerized implementations of these algorithms (GAP™, BESTFIT™, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, in addition to a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN™, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLocks IMProved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN™, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), FRAMEALIGN™, FRAMESEARCH™, DYNAMIC™, FILTER™, FSAP™ (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL™, GIBBS™, GENQUEST™, ISSC™ (Sensitive Sequence Comparison), LALIGN™ (Local Sequence Alignment), LCP™ (Local Content Program), MACAW™ (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP™, MBLKN™, PIMA™ (Pattern-Induced Multi-sequence Alignment), SAGA™ (Sequence Alignment by Genetic Algorithm) and WHAT-IF™. Such alignment programs can also be used to screen genome databases to identify polynucleotide sequences having substantially identical sequences. A number of genome databases are available, for example, a substantial portion of the human genome is available as part of the Human Genome Sequencing Project (Gibbs, 1995). At least twenty-one other genomes have already been sequenced, including, for example, *M. genitalium* (Fraser *et al.*, 1995), M. jannaschii (Bult *et al.*, 1996), *H. influenzae* (Fleischmann *et al.*, 1995), *E. coli* (Blattner *et al.*, 1997) and yeast (*S. cerevisiae*) (Mewes *et al.*, 1997) and *D. melanogaster* (Adams *et al.*, 2000). Significant progress has also been made in sequencing the genomes of model organism, such as mouse, *C. elegans* and *Arabadopsis sp.* Several databases containing genomic information annotated with some functional information are maintained by different organizations and may be accessible via the internet.

One example of a useful algorithm is BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, Nuc. Acids Res. 25:3389-3402, 1977 and Altschul *et al.*, J. Mol. Biol. 215:403-410, 1990, respectively. Software for performing BLAST analyses

is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3 and expectations (E) of 10 and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N= -4 and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, *e.g.*, Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873, 1993). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more in one aspect less than about 0.01 and most in one aspect less than about 0.001.

In one aspect, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is in one aspect obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are in one aspect identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. In one aspect, the scoring matrix used is the BLOSUM62 matrix (Gonnet (1992) Science 256:1443-1445; Henikoff and Henikoff (1993) Proteins 17:49-61). Less in one aspect, the PAM or PAM250 matrices may also be used (see, *e.g.*, Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). BLAST programs are accessible through the U.S. National Library of Medicine.

The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some aspects, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

Computer systems and computer program products

To determine and identify sequence identities, structural homologies, motifs and the like in silico, a nucleic acid or polypeptide sequence of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer.

Accordingly, the invention provides computers, computer systems, computer readable mediums, computer programs products and the like recorded or stored thereon the nucleic acid and polypeptide sequences of the invention. As used herein, the words

"recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid and/or polypeptide sequences of the invention.

The polypeptides of the invention include the polypeptide sequences of the invention, e.g., the exemplary sequences of the invention, and sequences substantially identical thereto, and fragments of any of the preceding sequences. Substantially identical, or homologous, polypeptide sequences refer to a polypeptide sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity (homology) to an exemplary sequence of the invention.

Homology (sequence identity) may be determined using any of the computer programs and parameters described herein. A nucleic acid or polypeptide sequence of the invention can be stored, recorded and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid sequences of the invention, one or more of the polypeptide sequences of the invention. Another aspect of the invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, or 20 or more nucleic acid or polypeptide sequences of the invention.

Another aspect of the invention is a computer readable medium having recorded thereon one or more of the nucleic acid sequences of the invention. Another aspect of the invention is a computer readable medium having recorded thereon one or more of the polypeptide sequences of the invention. Another aspect of the invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, or 20 or more of the nucleic acid or polypeptide sequences as set forth above.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Aspects of the invention include systems (*e.g.*, internet based systems), particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 1. As used herein, "a computer system" refers to the hardware components, software components and data storage components used to analyze a nucleotide sequence of a nucleic acid sequence of the invention, or a polypeptide sequence of the invention. The computer system 100 typically includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as, for example, the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq, AMD or International Business Machines.

Typically the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular aspect, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (in one aspect implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some aspects, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, or a modem capable of connection to a remote data storage system (*e.g.*, via internet) etc. In some aspects, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, (such as search tools, compare tools and modeling tools etc.) may reside in main memory 115 during execution.

In some aspects, the computer system 100 may further comprise a sequence comparison algorithm for comparing a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, stored on a computer readable medium to a reference nucleotide or polypeptide sequence(s) stored on a computer readable medium. A "sequence comparison algorithm" refers to one or more programs which are implemented (locally or remotely) on the computer system 100 to compare a nucleotide sequence with other nucleotide sequences and/or compounds stored within a data storage means. For example, the sequence comparison algorithm may compare the nucleotide sequences of a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies or structural motifs.

Figure 2 is a flow diagram illustrating one aspect of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The

parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to a nucleic acid sequence of the invention, or a polypeptide sequence of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the above described nucleic acid code a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some aspects, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30 or 40 or more of the nucleic acid sequences of the invention, or the polypeptide sequences of the invention.

Another aspect of the invention is a method for determining the level of homology between a nucleic acid sequence of the invention, or a polypeptide sequence of the invention and a reference nucleotide sequence. The method including reading the nucleic acid code or the polypeptide code and the reference nucleotide or polypeptide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code or polypeptide code and the reference nucleotide or polypeptide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, (*e.g.*, BLAST2N with the default parameters or with any modified parameters). The method may be implemented using the computer systems described above. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30 or 40 or more of the above described nucleic acid sequences of the invention, or the polypeptide sequences of the invention through use of the computer program and determining homology between the nucleic acid codes or polypeptide codes and reference nucleotide sequences or polypeptide sequences.

Figure 3 is a flow diagram illustrating one aspect of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it is in one aspect in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there are not any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is

displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of a nucleic acid sequence as set forth in the invention, to one or more reference nucleotide sequences in order to determine whether the nucleic acid code of the invention, differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or a nucleic acid sequence of the invention. In one aspect, the computer program may be a program which determines whether a nucleic acid sequence of the invention, contains a single nucleotide polymorphism (SNP) with respect to a reference nucleotide sequence.

Accordingly, another aspect of the invention is a method for determining whether a nucleic acid sequence of the invention, differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some aspects, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 3. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 40 or more of the nucleic acid sequences of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other aspects the computer based system may further comprise an identifier for identifying features within a nucleic acid sequence of the invention or a polypeptide sequence of the invention.

An "identifier" refers to one or more programs which identifies certain features within a nucleic acid sequence of the invention, or a polypeptide sequence of the invention. In one aspect, the identifier may comprise a program which identifies an open reading frame in a nucleic acid sequence of the invention.

Figure 4 is a flow diagram illustrating one aspect of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group. Alternatively, the features may be structural polypeptide motifs such as alpha helices, beta sheets, or functional polypeptide motifs such as enzymatic active sites, helix-turn-helix motifs or other motifs known to those skilled in the art.

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence. It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

Accordingly, another aspect of the invention is a method of identifying a feature within a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, comprising reading the nucleic acid code(s) or polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) with the computer program. In one aspect, computer program comprises a computer program which identifies open reading frames. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30,

or 40 of the nucleic acid sequences of the invention, or the polypeptide sequences of the invention, through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

A nucleic acid sequence of the invention, or a polypeptide sequence of the invention, may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, a nucleic acid sequence of the invention, or a polypeptide sequence of the invention, may be stored as text in a word processing file, such as Microsoft WORD™ or WORDPERFECT™ or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2™, SYBASE™, or ORACLE™. In addition, many computer programs and databases may be used as sequence comparison algorithms, identifiers, or sources of reference nucleotide sequences or polypeptide sequences to be compared to a nucleic acid sequence of the invention, or a polypeptide sequence of the invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid sequences of the invention, or the polypeptide sequences of the invention.

The programs and databases which may be used include, but are not limited to: MACPATTERN™ (EMBL), DISCOVERYBASE™ (Molecular Applications Group), GENEMINE™ (Molecular Applications Group), LOOK™ (Molecular Applications Group), MACLOOK™ (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, J. Mol. Biol. 215: 403, 1990), FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB (Brutlag *et al.* Comp. App. Biosci. 6:237-245, 1990), CATALYST™ (Molecular Simulations Inc.), CATALYST™/SHAPE™ (Molecular Simulations Inc.), CERIUS$^2$.DBACCESS™ (Molecular Simulations Inc.), HYPOGEN™ (Molecular Simulations Inc.), INSIGHT II™, (Molecular Simulations Inc.), DISCOVER™ (Molecular Simulations Inc.), CHARMm™ (Molecular Simulations Inc.), FELIX™ (Molecular Simulations Inc.), DELPHI™, (Molecular Simulations Inc.), QUANTEMM™, (Molecular Simulations Inc.), HOMOLOGY™ (Molecular Simulations Inc.), MODELER™ (Molecular Simulations Inc.), ISIS™ (Molecular Simulations Inc.), QUANTA™/Protein Design (Molecular Simulations Inc.), WEBLAB™ (Molecular Simulations Inc.), WEBLAB DIVERSITY EXPLORER™ (Molecular Simulations Inc.), GENE EXPLORER™ (Molecular Simulations Inc.), SEQFOLD™ (Molecular Simulations Inc.), the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents' World Drug Index database, the BioByteMasterFile database, the Genbank database and the Genseqn database.

Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites and enzymatic cleavage sites.

Hybridization of nucleic acids

The invention provides isolated, synthetic or recombinant nucleic acids that hybridize under stringent conditions to an exemplary sequence of the invention (e.g., SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9). The stringent conditions can be highly stringent conditions, medium stringent conditions and/or low stringent conditions, including the high and reduced stringency conditions described herein. In one aspect, it is the stringency of the wash conditions that set forth the conditions which determine whether a nucleic acid is within the scope of the invention, as discussed below.

"Hybridization" refers to the process by which a nucleic acid strand joins with a complementary strand through base pairing. Hybridization reactions can be sensitive and selective so that a particular sequence of interest can be identified even in samples in which it is present at low concentrations. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature. In alternative aspects, nucleic acids of the invention are defined by their ability to hybridize under various stringency conditions (e.g., high, medium, and low), as set forth herein.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In one aspect, hybridization occurs under high stringency conditions, e.g., at 42°C in 50% formamide, 5X SSPE, 0.3% SDS and 200 μg /ml sheared and denatured salmon sperm DNA. Hybridization could occur under these reduced stringency conditions, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to

pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

In alternative aspects, nucleic acids of the invention as defined by their ability to hybridize under stringent conditions can be between about five residues and the full length of nucleic acid of the invention; e.g., they can be at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, or more, residues in length. Nucleic acids shorter than full length are also included. These nucleic acids can be useful as, e.g., hybridization probes, labeling probes, PCR oligonucleotide probes, iRNA (single or double stranded), antisense or sequences encoding antibody binding peptides (epitopes), motifs, active sites and the like.

In one aspect, nucleic acids of the invention are defined by their ability to hybridize under high stringency comprises conditions of about 50% formamide at about 37°C to 42°C. In one aspect, nucleic acids of the invention are defined by their ability to hybridize under reduced stringency comprising conditions in about 35% to 25% formamide at about 30°C to 35°C.

Alternatively, nucleic acids of the invention are defined by their ability to hybridize under high stringency comprising conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and a repetitive sequence blocking nucleic acid, such as cot-1 or salmon sperm DNA (e.g., 200 μg /ml sheared and denatured salmon sperm DNA). In one aspect, nucleic acids of the invention are defined by their ability to hybridize under reduced stringency conditions comprising 35% formamide at a reduced temperature of 35°C.

In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (*e.g.*, GC v. AT content) and nucleic acid type (*e.g.*, RNA v. DNA) of the hybridizing regions of the nucleic acids can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter.

Hybridization may be carried out under conditions of low stringency, moderate stringency or high stringency. As an example of nucleic acid hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM $NaH_2PO_4$, pH 7.0, 5.0

mM $Na_2EDTA$, 0.5% SDS, 10X Denhardt's and 0.5 mg/ml polyriboadenylic acid. Approximately 2 X $10^7$ cpm (specific activity 4-9 X $10^8$ cpm/ug) of $^{32}P$ end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM $Na_2EDTA$) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at $T_m$-10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

All of the foregoing hybridizations would be considered to be under conditions of high stringency.

Following hybridization, a filter can be washed to remove any non-specifically bound detectable probe. The stringency used to wash the filters can also be varied depending on the nature of the nucleic acids being hybridized, the length of the nucleic acids being hybridized, the degree of complementarity, the nucleotide sequence composition (*e.g.*, GC v. AT content) and the nucleic acid type (*e.g.*, RNA v. DNA). Examples of progressively higher stringency condition washes are as follows: 2X SSC, 0.1% SDS at room temperature for 15 minutes (low stringency); 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour (moderate stringency); 0.1X SSC, 0.5% SDS for 15 to 30 minutes at between the hybridization temperature and 68°C (high stringency); and 0.15M NaCl for 15 minutes at 72°C (very high stringency). A final low stringency wash can be conducted in 0.1X SSC at room temperature. The examples above are merely illustrative of one set of conditions that can be used to wash filters. One of skill in the art would know that there are numerous recipes for different stringency washes. Some other examples are given below.

In one aspect, hybridization conditions comprise a wash step comprising a wash for 30 minutes at room temperature in a solution comprising 1X 150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM $Na_2EDTA$, 0.5% SDS, followed by a 30 minute wash in fresh solution.

Nucleic acids which have hybridized to the probe are identified by autoradiography or other conventional techniques.

The above procedure may be modified to identify nucleic acids having decreasing levels of homology to the probe sequence. For example, to obtain nucleic acids of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to

42°C in a hybridization buffer having a Na+ concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C. A specific example of "moderate" hybridization conditions is when the above hybridization is conducted at 55°C. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 45°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide. A specific example of "moderate" hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 10% formamide.

However, the selection of a hybridization format is not critical - it is the stringency of the wash conditions that set forth the conditions which determine whether a nucleic acid is within the scope of the invention. Wash conditions used to identify nucleic acids within the scope of the invention include, e.g.: a salt concentration of about 0.02 molar at pH 7 and a temperature of at least about 50°C or about 55°C to about 60°C; or, a salt concentration of about 0.15 M NaCl at 72°C for about 15 minutes; or, a salt concentration of about 0.2X SSC at a temperature of at least about 50°C or about 55°C to about 60°C for about 15 to about 20 minutes; or, the hybridization complex is washed twice with a solution with a salt concentration of about 2X SSC containing 0.1% SDS at room temperature for 15 minutes and then washed twice by 0.1X SSC containing 0.1% SDS at 68oC for 15 minutes; or, equivalent conditions. See Sambrook, Tijssen and Ausubel for a description of SSC buffer and equivalent conditions.

These methods may be used to isolate nucleic acids of the invention. For example, the preceding methods may be used to isolate nucleic acids having a sequence with at least about 97%, at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, or at least 50% sequence identity (homology) to a nucleic acid sequence selected from the group consisting of one of the sequences of the invention, or fragments comprising at least about 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases thereof and the sequences

complementary thereto. Sequence identity (homology) may be measured using the alignment algorithm. For example, the homologous polynucleotides may have a coding sequence which is a naturally occurring allelic variant of one of the coding sequences described herein. Such allelic variants may have a substitution, deletion or addition of one or more nucleotides when compared to the nucleic acids of the invention. Additionally, the above procedures may be used to isolate nucleic acids which encode polypeptides having at least about 99%, 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, or at least 50% sequence identity (homology) to a polypeptide of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof as determined using a sequence alignment algorithm (*e.g.*, such as the FASTA version 3.0t78 algorithm with the default parameters).

Oligonucleotides probes and methods for using them

The invention also provides nucleic acid probes that can be used, e.g., for identifying nucleic acids encoding a polypeptide with an enzyme, structural or binding activity or fragments thereof or for identifying polypeptide, enzyme, protein, e.g. structural or binding protein, genes. In one aspect, the probe comprises at least 10 consecutive bases of a nucleic acid of the invention. Alternatively, a probe of the invention can be at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 150 or about 10 to 50, about 20 to 60 about 30 to 70, consecutive bases of a sequence as set forth in a nucleic acid of the invention. The probes identify a nucleic acid by binding and/or hybridization. The probes can be used in arrays of the invention, see discussion below, including, e.g., capillary arrays. The probes of the invention can also be used to isolate other nucleic acids or polypeptides.

The isolated nucleic acids of the invention, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of the invention, or the sequences complementary thereto may also be used as probes to determine whether a biological sample, such as a soil sample, contains an organism having a nucleic acid sequence of the invention or an organism from which the nucleic acid was obtained. In such procedures, a biological sample potentially harboring the organism from which the nucleic acid was isolated is obtained and nucleic acids are obtained from the sample. The nucleic acids are

contacted with the probe under conditions which permit the probe to specifically hybridize to any complementary sequences from which are present therein.

Where necessary, conditions which permit the probe to specifically hybridize to complementary sequences may be determined by placing the probe in contact with complementary sequences from samples known to contain the complementary sequence as well as control sequences which do not contain the complementary sequence. Hybridization conditions, such as the salt concentration of the hybridization buffer, the formamide concentration of the hybridization buffer, or the hybridization temperature, may be varied to identify conditions which allow the probe to hybridize specifically to complementary nucleic acids.

If the sample contains the organism from which the nucleic acid was isolated, specific hybridization of the probe is then detected. Hybridization may be detected by labeling the probe with a detectable agent such as a radioactive isotope, a fluorescent dye or an enzyme capable of catalyzing the formation of a detectable product.

Many methods for using the labeled probes to detect the presence of complementary nucleic acids in a sample are familiar to those skilled in the art. These include Southern Blots, Northern Blots, colony hybridization procedures and dot blots. Protocols for each of these procedures are provided in Ausubel *et al.* Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. (1997) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory Press (1989.

Alternatively, more than one probe (at least one of which is capable of specifically hybridizing to any complementary sequences which are present in the nucleic acid sample), may be used in an amplification reaction to determine whether the sample contains an organism containing a nucleic acid sequence of the invention (*e.g.*, an organism from which the nucleic acid was isolated). Typically, the probes comprise oligonucleotides. In one aspect, the amplification reaction may comprise a PCR reaction. PCR protocols are described in Ausubel and Sambrook, *supra.* Alternatively, the amplification may comprise a ligase chain reaction, 3SR, or strand displacement reaction. (See Barany, F., "The Ligase Chain Reaction in a PCR World", *PCR Methods and Applications* 1:5-16, 1991; E. Fahy *et al.*, "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", *PCR Methods and Applications* 1:25-33, 1991; and Walker G.T. *et al.*, "Strand Displacement Amplification-an Isothermal *in vitro* DNA Amplification Technique", *Nucleic Acid Research* 20:1691-1696, 1992). In such procedures, the nucleic acids in the sample are

contacted with the probes, the amplification reaction is performed and any resulting amplification product is detected. The amplification product may be detected by performing gel electrophoresis on the reaction products and staining the gel with an intercalator such as ethidium bromide. Alternatively, one or more of the probes may be labeled with a radioactive isotope and the presence of a radioactive amplification product may be detected by autoradiography after gel electrophoresis.

Probes derived from sequences near the ends of the sequences of the invention, may also be used in chromosome walking procedures to identify clones containing genomic sequences located adjacent to the sequences of the invention. Such methods allow the isolation of genes which encode additional proteins from the host organism.

The isolated nucleic acids of the invention, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of the invention, or the sequences complementary thereto may be used as probes to identify and isolate related nucleic acids. In some aspects, the related nucleic acids may be cDNAs or genomic DNAs from organisms other than the one from which the nucleic acid was isolated. For example, the other organisms may be related organisms. In such procedures, a nucleic acid sample is contacted with the probe under conditions which permit the probe to specifically hybridize to related sequences. Hybridization of the probe to nucleic acids from the related organism is then detected using any of the methods described above.

By varying the stringency of the hybridization conditions used to identify nucleic acids, such as cDNAs or genomic DNAs, which hybridize to the detectable probe, nucleic acids having different levels of homology to the probe can be identified and isolated. Stringency may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature, $T_m$, is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly complementary probe. Very stringent conditions are selected to be equal to or about 5°C lower than the $T_m$ for a particular probe. The melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature ($T_m$) is calculated using the formula: $T_m=81.5+16.6(\log [Na+])+0.41(\text{fraction G+C})-(600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation: $T_m$=81.5+16.6(log [Na+])+0.41(fraction G+C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg/ml denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg/ml denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra.*

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the $T_m$. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 5-10°C below the $T_m$. In one aspect, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Usually, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

Inhibiting Expression of Polypeptides, Enzymes, Proteins

The invention provides nucleic acids complementary to (e.g., antisense sequences to) the nucleic acids of the invention, e.g., nucleic acids comprising antisense, iRNA, ribozymes. Nucleic acids of the invention comprising antisense sequences can be capable of inhibiting the transport, splicing or transcription of polypeptide, enzyme, protein, e.g. structural or binding protein genes. The inhibition can be effected through the targeting of genomic DNA or messenger RNA. The transcription or function of targeted nucleic acid can be inhibited, for example, by hybridization and/or cleavage. In one aspect, inhibitors of the invention include oligonucleotides which are able to either bind a polypeptide, enzyme, protein, e.g. structural or binding protein, gene or message, in either case preventing or inhibiting the production or function of a polypeptide, enzyme, protein, e.g. structural or binding protein. The association can be through sequence specific hybridization. Another useful class of inhibitors includes oligonucleotides which cause inactivation or cleavage of a polypeptide, enzyme, protein, e.g. structural or binding protein, message. The oligonucleotide of the invention can have enzyme activity, binding activity and/or structural activity to cause cleavage, e.g., of a message (transcript), for

example, oligonucleotides of the invention can comprise or consist of a ribozyme, an RNAi, and antisense oligo, and the like. The oligonucleotide can be chemically modified or conjugated to an enzyme or composition capable of cleaving the complementary nucleic acid. A pool of many different such oligonucleotides can be screened for those with the desired activity. Thus, the invention provides various compositions for the inhibition of a polypeptide, enzyme, protein, e.g. structural or binding protein, expression on a nucleic acid and/or protein level, e.g., antisense, iRNA and ribozymes comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, sequences of the invention and the anti-polypeptide, anti-enzyme, anti-protein, e.g. anti-structural or anti-binding protein antibodies of the invention.

Inhibition of a polypeptide, enzyme, protein, e.g. structural or binding protein, expression can have a variety of industrial applications. For example, inhibition of a polypeptide, enzyme, protein, e.g. structural or binding protein, expression can slow or prevent spoilage. In one aspect, use of compositions of the invention that inhibit the expression and/or activity of a polypeptide, enzyme, protein, e.g. structural or binding protein, e.g., antibodies, antisense oligonucleotides, ribozymes and RNAi, are used to slow or prevent spoilage. Thus, in one aspect, the invention provides methods and compositions comprising application onto a plant or plant product (e.g., a cereal, a grain, a fruit, seed, root, leaf, etc.) antibodies, antisense oligonucleotides, ribozymes and RNAi of the invention to slow or prevent spoilage. These compositions also can be expressed by the plant (e.g., a transgenic plant) or another organism (e.g., a bacterium or other microorganism transformed with a polypeptide, enzyme, protein, e.g. structural or binding protein, gene of the invention).

The compositions of the invention for the inhibition of a polypeptide, enzyme, protein, e.g. structural or binding protein, expression, e.g., antisense, iRNA (e.g., siRNA, miRNA), ribozymes, antibodies, can be used as pharmaceutical compositions, e.g., as anti-pathogen agents or in other therapies, e.g., as anti-microbials for, e.g., *Salmonella*, or to neutralize a biological warfare agent, e.g., anthrax.

*Antisense Oligonucleotides*

The invention provides antisense oligonucleotides capable of binding a polypeptide, enzyme, protein, e.g. structural or binding protein, message which, in one aspect, can inhibit a polypeptide, enzyme, protein, e.g. structural or binding protein, activity by targeting mRNA. Strategies for designing antisense oligonucleotides are well described in the scientific and patent literature, and the skilled artisan can design such a

polypeptide, enzyme, protein, e.g. structural or binding protein, oligonucleotides using the novel reagents of the invention. For example, gene walking/ RNA mapping protocols to screen for effective antisense oligonucleotides are well known in the art, see, e.g., Ho (2000) Methods Enzymol. 314:168-183, describing an RNA mapping assay, which is based on standard molecular techniques to provide an easy and reliable method for potent antisense sequence selection. See also Smith (2000) Eur. J. Pharm. Sci. 11:191-198.

Naturally occurring nucleic acids are used as antisense oligonucleotides. The antisense oligonucleotides can be of any length; for example, in alternative aspects, the antisense oligonucleotides are between about 5 to 100, about 10 to 80, about 15 to 60, about 18 to 40. The optimal length can be determined by routine screening. The antisense oligonucleotides can be present at any concentration. The optimal concentration can be determined by routine screening. A wide variety of synthetic, non-naturally occurring nucleotide and nucleic acid analogues are known which can address this potential problem. For example, peptide nucleic acids (PNAs) containing non-ionic backbones, such as N-(2-aminoethyl) glycine units can be used. Antisense oligonucleotides having phosphorothioate linkages can also be used, as described in WO 97/03211; WO 96/39154; Mata (1997) Toxicol Appl Pharmacol 144:189-197; Antisense Therapeutics, ed. Agrawal (Humana Press, Totowa, N.J., 1996). Antisense oligonucleotides having synthetic DNA backbone analogues provided by the invention can also include phosphoro-dithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, and morpholino carbamate nucleic acids, as described above.

Combinatorial chemistry methodology can be used to create vast numbers of oligonucleotides that can be rapidly screened for specific oligonucleotides that have appropriate binding affinities and specificities toward any target, such as the sense and antisense a polypeptide, enzyme, protein, e.g. structural or binding protein, sequences of the invention (see, e.g., Gold (1995) J. of Biol. Chem. 270:13581-13584).

*Inhibitory Ribozymes*

The invention provides ribozymes capable of binding a polypeptide, enzyme, protein, e.g. structural or binding protein, message. These ribozymes can inhibit a polypeptide, enzyme, protein, e.g. structural or binding protein, activity by, e.g., targeting mRNA. Strategies for designing ribozymes and selecting the polypeptide, enzyme, protein, e.g. structural or binding protein-specific antisense sequence for targeting are well described in the scientific and patent literature, and the skilled artisan

can design such ribozymes using the novel reagents of the invention. Ribozymes act by binding to a target RNA through the target RNA binding portion of a ribozyme which is held in close proximity to an enzymatic portion of the RNA that cleaves the target RNA. Thus, the ribozyme recognizes and binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cleave and inactivate the target RNA. Cleavage of a target RNA in such a manner will destroy its ability to direct synthesis of an encoded protein if the cleavage occurs in the coding sequence. After a ribozyme has bound and cleaved its RNA target, it can be released from that RNA to bind and cleave new targets repeatedly.

In some circumstances, the enzymatic nature of a ribozyme can be advantageous over other technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its transcription, translation or association with another molecule) as the effective concentration of ribozyme necessary to effect a therapeutic treatment can be lower than that of an antisense oligonucleotide. This potential advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, a ribozyme is typically a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding, but also on the mechanism by which the molecule inhibits the expression of the RNA to which it binds. That is, the inhibition is caused by cleavage of the RNA target and so specificity is defined as the ratio of the rate of cleavage of the targeted RNA over the rate of cleavage of non-targeted RNA. This cleavage mechanism is dependent upon factors additional to those involved in base pairing. Thus, the specificity of action of a ribozyme can be greater than that of antisense oligonucleotide binding the same RNA site.

The ribozyme of the invention, e.g., an enzymatic ribozyme RNA molecule, can be formed in a hammerhead motif, a hairpin motif, as a hepatitis delta virus motif, a group I intron motif and/or an RNaseP-like RNA in association with an RNA guide sequence. Examples of hammerhead motifs are described by, e.g., Rossi (1992) Aids Research and Human Retroviruses 8:183; hairpin motifs by Hampel (1989) Biochemistry 28:4929, and Hampel (1990) Nuc. Acids Res. 18:299; the hepatitis delta virus motif by Perrotta (1992) Biochemistry 31:16; the RNaseP motif by Guerrier-Takada (1983) Cell 35:849; and the group I intron by Cech U.S. Pat. No. 4,987,071. The recitation of these specific motifs is not intended to be limiting. Those skilled in the art will recognize that a ribozyme of the invention, e.g., an enzymatic RNA molecule of this

invention, can have a specific substrate binding site complementary to one or more of the target gene RNA regions. A ribozyme of the invention can have a nucleotide sequence within or surrounding that substrate binding site which imparts an RNA cleaving activity to the molecule.

*RNA interference (RNAi)*

In one aspect, the invention provides an RNA inhibitory molecule, a so-called "RNAi" molecule, comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, sequence of the invention. The RNAi molecule comprises a double-stranded RNA (dsRNA) molecule. The RNAi can inhibit expression of a polypeptide, enzyme, protein, e.g. structural or binding protein, gene. In one aspect, the RNAi is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length. While the invention is not limited by any particular mechanism of action, the RNAi can enter a cell and cause the degradation of a single-stranded RNA (ssRNA) of similar or identical sequences, including endogenous mRNAs. When a cell is exposed to double-stranded RNA (dsRNA), mRNA from the homologous gene is selectively degraded by a process called RNA interference (RNAi). A possible basic mechanism behind RNAi, e.g., siRNA for inhibiting transcription and/or miRNA to inhibit translation, is the breaking of a double-stranded RNA (dsRNA) matching a specific gene sequence into short pieces called short interfering RNA, which trigger the degradation of mRNA that matches its sequence. In one aspect, the RNAi's of the invention are used in gene-silencing therapeutics, see, e.g., Shuey (2002) Drug Discov. Today 7:1040-1046. In one aspect, the invention provides methods to selectively degrade RNA using the RNAi's of the invention. The process may be practiced *in vitro*, *ex vivo* or *in vivo*. In one aspect, the RNAi molecules of the invention can be used to generate a loss-of-function mutation in a cell, an organ or an animal. Methods for making and using RNAi molecules for selectively degrade RNA are well known in the art, see, e.g., U.S. Patent No. 6,506,559; 6,511,824; 6,515,109; 6,489,127.

Modification of Nucleic Acids

The invention provides methods of generating variants of the nucleic acids of the invention, e.g., those encoding a polypeptide, enzyme, protein, e.g. structural or binding protein. These methods can be repeated or used in various combinations to generate a polypeptide, enzyme, protein, e.g. structural or binding protein, having an altered or different activity or an altered or different stability from that of a polypeptide, enzyme,

protein, e.g. structural or binding protein, encoded by the template nucleic acid. These methods also can be repeated or used in various combinations, e.g., to generate variations in gene/ message expression, message translation or message stability. In another aspect, the genetic composition of a cell is altered by, e.g., modification of a homologous gene ex vivo, followed by its reinsertion into the cell.

A nucleic acid of the invention can be altered by any means. For example, random or stochastic methods, or, non-stochastic, or "directed evolution," methods, see, e.g., U.S. Patent No. 6,361,974. Methods for random mutation of genes are well known in the art, see, e.g., U.S. Patent No. 5,830,696. For example, mutagens can be used to randomly mutate a gene. Mutagens include, e.g., ultraviolet light or gamma irradiation, or a chemical mutagen, e.g., mitomycin, nitrous acid, photoactivated psoralens, alone or in combination, to induce DNA breaks amenable to repair by recombination. Other chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. Other mutagens are analogues of nucleotide precursors, e.g., nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine. These agents can be added to a PCR reaction in place of the nucleotide precursor thereby mutating the sequence. Intercalating agents such as proflavine, acriflavine, quinacrine and the like can also be used.

Any technique in molecular biology can be used, e.g., random PCR mutagenesis, see, e.g., Rice (1992) Proc. Natl. Acad. Sci. USA 89:5467-5471; or, combinatorial multiple cassette mutagenesis, see, e.g., Crameri (1995) Biotechniques 18:194-196. Alternatively, nucleic acids, e.g., genes, can be reassembled after random, or "stochastic," fragmentation, see, e.g., U.S. Patent Nos. 6,291,242; 6,287,862; 6,287,861; 5,955,358; 5,830,721; 5,824,514; 5,811,238; 5,605,793. In alternative aspects, modifications, additions or deletions are introduced by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR), recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification

mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation, and/or a combination of these and other methods.

The following publications describe a variety of recursive recombination procedures and/or methods which can be incorporated into the methods of the invention: Stemmer (1999) "Molecular breeding of viruses for targeting and other clinical properties" Tumor Targeting 4:1-4; Ness (1999) Nature Biotechnology 17:893-896; Chang (1999) "Evolution of a cytokine using DNA family shuffling" Nature Biotechnology 17:793-797; Minshull (1999) "Protein evolution by molecular breeding" Current Opinion in Chemical Biology 3:284-290; Christians (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" Nature Biotechnology 17:259-264; Crameri (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288-291; Crameri (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" Proc. Natl. Acad. Sci. USA 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" Current Opinion in Biotechnology 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor `headpiece dimer`" Journal of Molecular Biology 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" BioTechniques 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gene and entire plasmid form large numbers of oligodeoxyribonucleotides" Gene, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" Science 270: 1510; Stemmer (1995) "Searching Sequence Space" Bio/Technology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." Proc. Natl. Acad. Sci. USA 91:10747-10751.

Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale et al. (1996) "Oligonucleotide-directed random

mutagenesis using the phosphorothioate method" Methods Mol. Biol. 57:369-374; Smith (1985) "In vitro mutagenesis" Ann. Rev. Genet. 19:423-462; Botstein & Shortle (1985) "Strategies and applications of in vitro mutagenesis" Science 229:1193-1201; Carter (1986) "Site-directed mutagenesis" Biochem. J. 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Methods in Enzymol. 154, 367-382; and Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" Science 242:240-245); oligonucleotide-directed mutagenesis (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" Nucleic Acids Res. 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" Methods in Enzymol. 100:468-500; and Zoller (1987) Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" Methods in Enzymol. 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" Nucl. Acids Res. 13: 8749-8764; Taylor (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucl. Acids Res. 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" Nucl. Acids Res. 16:791-802; and Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucl. Acids Res. 16: 803-814); mutagenesis using gapped duplex DNA (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to

oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; and Fritz (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*" Nucl. Acids Res. 16: 6987-6999).

Additional protocols that can be used to practice the invention include point mismatch repair (Kramer (1984) "Point Mismatch Repair" Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), restriction-selection and restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; and Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale `shot-gun` gene synthesis" Nucl. Acids Res. 13: 3305-3316), double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455. "Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Protocols that can be used to practice the invention are described, e.g., in U.S. Patent Nos. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for In Vitro Recombination;" U.S. Pat. No. 5,811,238 to Stemmer et al. (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer et al. (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, et al. (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat.

No. 5,837,458 to Minshull, et al. (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 by Stemmer and Crameri "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen et al. "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen et al. "Antigen Library Immunization;" WO 99/41369 by Punnonen et al. "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt et al., "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/27230 by Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection," WO 00/00632, "Methods for Generating Highly Diverse Libraries," WO 00/09679, "Methods for Obtaining in Vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences," WO 98/42832 by Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers," WO 99/29902 by Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences," WO 98/41653 by Vind, "An in Vitro Method for Construction of a DNA Library," WO 98/41622 by Borchert et al., "Method for Constructing a Library Using DNA Shuffling," and WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination."

Protocols that can be used to practice the invention (providing details regarding various diversity generating methods) are described, e.g., in U.S. Patent application serial no. (USSN) 09/407,800, "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed Sep. 28, 1999; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" by del Cardayre et al., United States Patent No. 6,379,964; "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., United States Patent Nos. 6,319,714;

6,368,861; 6,376,246; 6,423,542; 6,426,224 and PCT/US00/01203; "USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., United States Patent No. 6,436,675; "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed Jan. 18, 2000, (PCT/US00/01202) and, e.g. "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed Jul. 18, 2000 (U.S. Ser. No. 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" by Selifonov and Stemmer, filed Jan. 18, 2000 (PCT/US00/01138); and "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" by Affholter, filed Sep. 6, 2000 (U.S. Ser. No. 09/656,549); and United States Patent Nos. 6,177,263; 6,153,410.

Non-stochastic, or "directed evolution," methods include, e.g., saturation mutagenesis, such as Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR), or a combination thereof are used to modify the nucleic acids of the invention to generate a polypeptide, enzyme, protein, e.g. structural or binding protein, with new or altered properties (e.g., activity under highly acidic or alkaline conditions, high or low temperatures, and the like). Polypeptides encoded by the modified nucleic acids can be screened for an activity before testing for glucan hydrolysis or other activity. Any testing modality or protocol can be used, e.g., using a capillary array platform. See, e.g., U.S. Patent Nos. 6,361,974; 6,280,926; 5,939,250.

*Gene Site Saturation mutagenesis, or, GSSM*

The invention also provides methods for making enzyme using Gene Site Saturation mutagenesis, or, GSSM, as described herein, and also in U.S. Patent Nos. 6,171,820 and 6,579,258.

In one aspect, codon primers containing a degenerate N,N,G/T sequence are used to introduce point mutations into a polynucleotide, e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, or an antibody of the invention, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is represented at each amino acid position, e.g., an amino acid residue in an enzyme active site or ligand binding site targeted to be modified. These oligonucleotides can comprise a contiguous first homologous sequence, a degenerate N,N,G/T sequence, and, optionally, a

second homologous sequence. The downstream progeny translational products from the use of such oligonucleotides include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,G/T sequence includes codons for all 20 amino acids. In one aspect, one such degenerate oligonucleotide (comprised of, e.g., one degenerate N,N,G/T cassette) is used for subjecting each original codon in a parental polynucleotide template to a full range of codon substitutions. In another aspect, at least two degenerate cassettes are used – either in the same oligonucleotide or not, for subjecting at least two original codons in a parental polynucleotide template to a full range of codon substitutions. For example, more than one N,N,G/T sequence can be contained in one oligonucleotide to introduce amino acid mutations at more than one site. This plurality of N,N,G/T sequences can be directly contiguous, or separated by one or more additional nucleotide sequence(s). In another aspect, oligonucleotides serviceable for introducing additions and deletions can be used either alone or in combination with the codons containing an N,N,G/T sequence, to introduce any combination or permutation of amino acid additions, deletions, and/or substitutions.

In one aspect, simultaneous mutagenesis of two or more contiguous amino acid positions is done using an oligonucleotide that contains contiguous N,N,G/T triplets, i.e. a degenerate (N,N,G/T)n sequence. In another aspect, degenerate cassettes having less degeneracy than the N,N,G/T sequence are used. For example, it may be desirable in some instances to use (e.g. in an oligonucleotide) a degenerate triplet sequence comprised of only one N, where said N can be in the first second or third position of the triplet. Any other bases including any combinations and permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some instances to use (e.g. in an oligo) a degenerate N,N,N triplet sequence.

In one aspect, use of degenerate triplets (e.g., N,N,G/T triplets) allows for systematic and easy generation of a full range of possible natural amino acids (for a total of 20 amino acids) into each and every amino acid position in a polypeptide (in alternative aspects, the methods also include generation of less than all possible substitutions per amino acid residue, or codon, position). For example, for a 100 amino acid polypeptide, 2000 distinct species (i.e. 20 possible amino acids per position X 100 amino acid positions) can be generated. Through the use of an oligonucleotide or set of oligonucleotides containing a degenerate N,N,G/T triplet, 32 individual sequences can code for all 20 possible natural amino acids. Thus, in a reaction vessel in which a

parental polynucleotide sequence is subjected to saturation mutagenesis using at least one such oligonucleotide, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligonucleotide in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel. Nondegenerate oligonucleotides can optionally be used in combination with degenerate primers disclosed; for example, nondegenerate oligonucleotides can be used to generate specific point mutations in a working polynucleotide. This provides one means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes, and point mutations that cause the generation of stop codons and the corresponding expression of polypeptide fragments.

In one aspect, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, ) molecules such that all 20 natural amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide (other aspects use less than all 20 natural combinations). The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (e.g. cloned into a suitable host, e.g., *E. coli* host, using, e.g., an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in property (when compared to the parental polypeptide, such as increased glucan hydrolysis activity under alkaline or acidic conditions), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

In one aspect, upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid, and each of two favorable changes) and 3 positions. Thus, there are 3 x 3 x 3 or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (i.e. 2 at each of three positions) and no change at any position.

In yet another aspect, site-saturation mutagenesis can be used together with shuffling, chimerization, recombination and other mutagenizing processes, along with screening. This invention provides for the use of any mutagenizing process(es), including saturation mutagenesis, in an iterative manner. In one exemplification, the iterative use of any mutagenizing process(es) is used in combination with screening.

The invention also provides for the use of proprietary codon primers (containing a degenerate N,N,N sequence) to introduce point mutations into a polynucleotide, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is represented at each amino acid position; e.g., with Gene Site Saturation Mutagenesis (GSSM). The oligos used are comprised contiguously of a first homologous sequence, a degenerate N,N,N sequence and in one aspect but not necessarily a second homologous sequence. The downstream progeny translational products from the use of such oligos include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,N sequence includes codons for all 20 amino acids.

In one aspect, one such degenerate oligo (comprised of one degenerate N,N,N cassette) is used for subjecting each original codon in a parental polynucleotide template to a full range of codon substitutions. In another aspect, at least two degenerate N,N,N cassettes are used – either in the same oligo or not, for subjecting at least two original codons in a parental polynucleotide template to a full range of codon substitutions. Thus, more than one N,N,N sequence can be contained in one oligo to introduce amino acid mutations at more than one site. This plurality of N,N,N sequences can be directly contiguous, or separated by one or more additional nucleotide sequence(s). In another aspect, oligos serviceable for introducing additions and deletions can be used either alone or in combination with the codons containing an N,N,N sequence, to introduce any combination or permutation of amino acid additions, deletions and/or substitutions.

In a particular exemplification, it is possible to simultaneously mutagenize two or more contiguous amino acid positions using an oligo that contains contiguous N,N,N triplets, *i.e.* a degenerate $(N,N,N)_n$ sequence.

In another aspect, the present invention provides for the use of degenerate cassettes having less degeneracy than the N,N,N sequence. For example, it may be desirable in some instances to use (*e.g.* in an oligo) a degenerate triplet sequence comprised of only one N, where the N can be in the first second or third position of the

triplet. Any other bases including any combinations and permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some instances to use (*e.g.*, in an oligo) a degenerate N,N,N triplet sequence, N,N,G/T, or an N,N, G/C triplet sequence.

It is appreciated, however, that the use of a degenerate triplet (such as N,N,G/T or an N,N, G/C triplet sequence) as disclosed in the instant invention is advantageous for several reasons. In one aspect, this invention provides a means to systematically and fairly easily generate the substitution of the full range of possible amino acids (for a total of 20 amino acids) into each and every amino acid position in a polypeptide. Thus, for a 100 amino acid polypeptide, the invention provides a way to systematically and fairly easily generate 2000 distinct species (*i.e.*, 20 possible amino acids per position times 100 amino acid positions). It is appreciated that there is provided, through the use of an oligo containing a degenerate N,N,G/T or an N,N, G/C triplet sequence, 32 individual sequences that code for 20 possible amino acids. Thus, in a reaction vessel in which a parental polynucleotide sequence is subjected to saturation mutagenesis using one such oligo, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligo in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel.

This invention also provides for the use of nondegenerate oligos, which can optionally be used in combination with degenerate primers disclosed. It is appreciated that in some situations, it is advantageous to use nondegenerate oligos to generate specific point mutations in a working polynucleotide. This provides a means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes and point mutations that cause the generation of stop codons and the corresponding expression of polypeptide fragments.

Thus, in one aspect of this invention, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide molecules such that all 20 amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide. The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (*e.g.*, cloned into a suitable *E. coli* host using an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in

property (when compared to the parental polypeptide), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

It is appreciated that upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid and each of two favorable changes) and 3 positions. Thus, there are 3 x 3 x 3 or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (*i.e.*, 2 at each of three positions) and no change at any position.

Thus, in a non-limiting exemplification, this invention provides for the use of saturation mutagenesis in combination with additional mutagenization processes, such as process where two or more related polynucleotides are introduced into a suitable host cell such that a hybrid polynucleotide is generated by recombination and reductive reassortment.

In addition to performing mutagenesis along the entire sequence of a gene, the instant invention provides that mutagenesis can be use to replace each of any number of bases in a polynucleotide sequence, wherein the number of bases to be mutagenized is in one aspect every integer from 15 to 100,000. Thus, instead of mutagenizing every position along a molecule, one can subject every or a discrete number of bases (in one aspect a subset totaling from 15 to 100,000) to mutagenesis. In one aspect, a separate nucleotide is used for mutagenizing each position or group of positions along a polynucleotide sequence. A group of 3 positions to be mutagenized may be a codon. The mutations can be introduced using a mutagenic primer, containing a heterologous cassette, also referred to as a mutagenic cassette. Exemplary cassettes can have from 1 to 500 bases. Each nucleotide position in such heterologous cassettes be N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, or E, where E is any base that is not A, C, G, or T (E can be referred to as a designer oligo).

In a general sense, saturation mutagenesis is comprised of mutagenizing a complete set of mutagenic cassettes (wherein each cassette is in one aspect about 1-500 bases in length) in defined polynucleotide sequence to be mutagenized (wherein the sequence to be mutagenized is in one aspect from about 15 to 100,000 bases in length).

Thus, a group of mutations (ranging from 1 to 100 mutations) is introduced into each cassette to be mutagenized. A grouping of mutations to be introduced into one cassette can be different or the same from a second grouping of mutations to be introduced into a second cassette during the application of one round of saturation mutagenesis. Such groupings are exemplified by deletions, additions, groupings of particular codons and groupings of particular nucleotide cassettes.

Defined sequences to be mutagenized include a whole gene, pathway, cDNA, an entire open reading frame (ORF) and entire promoter, enhancer, repressor/transactivator, origin of replication, intron, operator, or any polynucleotide functional group. Generally, a "defined sequences" for this purpose may be any polynucleotide that a 15 base-polynucleotide sequence and polynucleotide sequences of lengths between 15 bases and 15,000 bases (this invention specifically names every integer in between). Considerations in choosing groupings of codons include types of amino acids encoded by a degenerate mutagenic cassette.

In one exemplification a grouping of mutations that can be introduced into a mutagenic cassette, this invention specifically provides for degenerate codon substitutions (using degenerate oligos) that code for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 amino acids at each position and a library of polypeptides encoded thereby.

*Synthetic Ligation Reassembly (SLR)*

The invention provides a non-stochastic gene modification system termed "synthetic ligation reassembly," or simply "SLR," a "directed evolution process," to generate polypeptides, e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, or antibodies of the invention, with new or altered properties. SLR is a method of ligating oligonucleotide fragments together non-stochastically. This method differs from stochastic oligonucleotide shuffling in that the nucleic acid building blocks are not shuffled, concatenated or chimerized randomly, but rather are assembled non-stochastically. See, e.g., U.S. Patent Nos. 6,773,900; 6,740,506; 6,713,282; 6,635,449; 6,605,449; 6,537,776.

In one aspect, SLR comprises the following steps: (a) providing a template polynucleotide, wherein the template polynucleotide comprises sequence encoding a homologous gene; (b) providing a plurality of building block polynucleotides, wherein the building block polynucleotides are designed to cross-over reassemble with the template polynucleotide at a predetermined sequence, and a building block

polynucleotide comprises a sequence that is a variant of the homologous gene and a sequence homologous to the template polynucleotide flanking the variant sequence; (c) combining a building block polynucleotide with a template polynucleotide such that the building block polynucleotide cross-over reassembles with the template polynucleotide to generate polynucleotides comprising homologous gene sequence variations.

SLR does not depend on the presence of high levels of homology between polynucleotides to be rearranged. Thus, this method can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over $10^{100}$ different chimeras. SLR can be used to generate libraries comprised of over $10^{1000}$ different progeny chimeras. Thus, aspects of the present invention include non-stochastic methods of producing a set of finalized chimeric nucleic acid molecule shaving an overall assembly order that is chosen by design. This method includes the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends. If more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). In one aspect, the annealed building pieces are treated with an enzyme, such as a ligase (e.g. T4 DNA ligase), to achieve covalent bonding of the building pieces.

In one aspect, the design of the oligonucleotide building blocks is obtained by analyzing a set of progenitor nucleic acid sequence templates that serve as a basis for producing a progeny set of finalized chimeric polynucleotides. These parental oligonucleotide templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, e.g., chimerized or shuffled. In one aspect of this method, the sequences of a plurality of parental nucleic acid templates are aligned in order to select one or more demarcation points. The demarcation points can be located at an area of homology, and are comprised of one or more nucleotides. These demarcation points are in one aspect shared by at least two of the progenitor templates. The demarcation points can thereby be used to delineate the

boundaries of oligonucleotide building blocks to be generated in order to rearrange the parental polynucleotides. The demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the final chimeric progeny molecules. A demarcation point can be an area of homology (comprised of at least one homologous nucleotide base) shared by at least two parental polynucleotide sequences. Alternatively, a demarcation point can be an area of homology that is shared by at least half of the parental polynucleotide sequences, or, it can be an area of homology that is shared by at least two thirds of the parental polynucleotide sequences. Even more in one aspect a serviceable demarcation points is an area of homology that is shared by at least three fourths of the parental polynucleotide sequences, or, it can be shared by at almost all of the parental polynucleotide sequences. In one aspect, a demarcation point is an area of homology that is shared by all of the parental polynucleotide sequences.

In one aspect, a ligation reassembly process is performed exhaustively in order to generate an exhaustive library of progeny chimeric polynucleotides. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, in another aspect, the assembly order (i.e. the order of assembly of each building block in the 5' to 3 sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic) as described above. Because of the non-stochastic nature of this invention, the possibility of unwanted side products is greatly reduced.

In another aspect, the ligation reassembly method is performed systematically. For example, the method is performed in order to generate a systematically compartmentalized library of progeny molecules, with compartments that can be screened systematically, e.g. one by one. In other words this invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, a design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, these methods allow a potentially very large number of progeny molecules to be examined systematically in smaller groups. Because of its ability to perform chimerizations in a manner that is highly flexible yet exhaustive and systematic as well, particularly when there is a low level of homology among the progenitor molecules, these methods provide for the generation of a library (or set) comprised of a

large number of progeny molecules. Because of the non-stochastic nature of the instant ligation reassembly invention, the progeny molecules generated in one aspect comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. The saturation mutagenesis and optimized directed evolution methods also can be used to generate different progeny molecular species. It is appreciated that the invention provides freedom of choice and control regarding the selection of demarcation points, the size and number of the nucleic acid building blocks, and the size and design of the couplings. It is appreciated, furthermore, that the requirement for intermolecular homology is highly relaxed for the operability of this invention. In fact, demarcation points can even be chosen in areas of little or no intermolecular homology. For example, because of codon wobble, i.e. the degeneracy of codons, nucleotide substitutions can be introduced into nucleic acid building blocks without altering the amino acid originally encoded in the corresponding progenitor template. Alternatively, a codon can be altered such that the coding for an originally amino acid is altered. This invention provides that such substitutions can be introduced into the nucleic acid building block in order to increase the incidence of intermolecular homologous demarcation points and thus to allow an increased number of couplings to be achieved among the building blocks, which in turn allows a greater number of progeny chimeric molecules to be generated.

In one aspect, the present invention provides a non-stochastic method termed synthetic gene reassembly, that is somewhat related to stochastic shuffling, save that the nucleic acid building blocks are not shuffled or concatenated or chimerized randomly, but rather are assembled non-stochastically.

The synthetic gene reassembly method does not depend on the presence of a high level of homology between polynucleotides to be shuffled. The invention can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over $10^{100}$ different chimeras. Conceivably, synthetic gene reassembly can even be used to generate libraries comprised of over $10^{1000}$ different progeny chimeras.

Thus, in one aspect, the invention provides a non-stochastic method of producing a set of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design, which method is comprised of the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, in one aspect, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends and, if more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). In a one aspect of the invention, the annealed building pieces are treated with an enzyme, such as a ligase (*e.g.*, T4 DNA ligase) to achieve covalent bonding of the building pieces.

In a another aspect, the design of nucleic acid building blocks is obtained upon analysis of the sequences of a set of progenitor nucleic acid templates that serve as a basis for producing a progeny set of finalized chimeric nucleic acid molecules. These progenitor nucleic acid templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, *i.e.* chimerized or shuffled.

In one exemplification, the invention provides for the chimerization of a family of related genes and their encoded family of related products. In a particular exemplification, the encoded products are enzymes. The polypeptide, enzyme, protein, e.g. structural or binding proteins of the present invention can be mutagenized in accordance with the methods described herein.

Thus according to one aspect of the invention, the sequences of a plurality of progenitor nucleic acid templates (*e.g.*, polynucleotides of the invention) are aligned in order to select one or more demarcation points, which demarcation points can be located at an area of homology. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks to be generated. Thus, the demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the progeny molecules.

Typically a serviceable demarcation point is an area of homology (comprised of at least one homologous nucleotide base) shared by at least two progenitor templates, but the demarcation point can be an area of homology that is shared by at least half of the progenitor templates, at least two thirds of the progenitor templates, at least three fourths of the progenitor templates and in one aspect at almost all of the progenitor templates. Even more in one aspect still a serviceable demarcation point is an area of homology that is shared by all of the progenitor templates.

In a one aspect, the gene reassembly process is performed exhaustively in order to generate an exhaustive library. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, the assembly order (*i.e.* the order of assembly of each building block in the 5' to 3 sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic). Because of the non-stochastic nature of the method, the possibility of unwanted side products is greatly reduced.

In another aspect, the method provides that the gene reassembly process is performed systematically, for example to generate a systematically compartmentalized library, with compartments that can be screened systematically, *e.g.*, one by one. In other words the invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, it allows a potentially very large number of progeny molecules to be examined systematically in smaller groups.

Because of its ability to perform chimerizations in a manner that is highly flexible yet exhaustive and systematic as well, particularly when there is a low level of homology among the progenitor molecules, the instant invention provides for the generation of a library (or set) comprised of a large number of progeny molecules. Because of the non-stochastic nature of the instant gene reassembly invention, the progeny molecules generated in one aspect comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. In a particularly aspect, such a generated library is comprised of greater than $10^3$ to greater than $10^{1000}$ different progeny molecular species.

In one aspect, a set of finalized chimeric nucleic acid molecules, produced as described is comprised of a polynucleotide encoding a polypeptide. According to one aspect, this polynucleotide is a gene, which may be a man-made gene. According to another aspect, this polynucleotide is a gene pathway, which may be a man-made gene pathway. The invention provides that one or more man-made genes generated by the invention may be incorporated into a man-made gene pathway, such as pathway operable in a eukaryotic organism (including a plant).

In another exemplification, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (*e.g.*, one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that can later be optionally removed in an *in vitro* process (*e.g.*, by mutagenesis) or in an *in vivo* process (*e.g.*, by utilizing the gene splicing ability of a host organism). It is appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.

Thus, according to another aspect, the invention provides that a nucleic acid building block can be used to introduce an intron. Thus, the invention provides that functional introns may be introduced into a man-made gene of the invention. The invention also provides that functional introns may be introduced into a man-made gene pathway of the invention. Accordingly, the invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced intron(s).

Accordingly, the invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially introduced intron(s). In one aspect, the artificially introduced intron(s) are functional in one or more host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing. The invention provides a process of producing man-made intron-containing polynucleotides to be introduced into host organisms for recombination and/or splicing.

A man-made gene produced using the invention can also serve as a substrate for recombination with another nucleic acid. Likewise, a man-made gene pathway produced using the invention can also serve as a substrate for recombination with another nucleic acid. In one aspect, the recombination is facilitated by, or occurs at, areas of homology between the man-made, intron-containing gene and a nucleic acid, which serves as a recombination partner. In one aspect, the recombination partner may also be a nucleic acid generated by the invention, including a man-made gene or a man-made gene pathway. Recombination may be facilitated by or may occur at areas of homology that exist at the one (or more) artificially introduced intron(s) in the man-made gene.

The synthetic gene reassembly method of the invention utilizes a plurality of nucleic acid building blocks, each of which in one aspect has two ligatable ends. The

two ligatable ends on each nucleic acid building block may be two blunt ends (*i.e.* each having an overhang of zero nucleotides), or in one aspect one blunt end and one overhang, or more in one aspect still two overhangs.

A useful overhang for this purpose may be a 3' overhang or a 5' overhang. Thus, a nucleic acid building block may have a 3' overhang or alternatively a 5' overhang or alternatively two 3' overhangs or alternatively two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design and is not random.

In one aspect, a nucleic acid building block is generated by chemical synthesis of two single-stranded nucleic acids (also referred to as single-stranded oligos) and contacting them so as to allow them to anneal to form a double-stranded nucleic acid building block.

A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large. Exemplary sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other exemplary size ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between) and upper limits of from 2 bp to 100, 000 bp (including every integer value in between).

Many methods exist by which a double-stranded nucleic acid building block can be generated that is serviceable for the invention; and these are known in the art and can be readily performed by the skilled artisan.

According to one aspect, a double-stranded nucleic acid building block is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another aspect, the two strands of a double-stranded nucleic acid building block are complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this aspect, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. In one aspect the codon degeneracy is introduced using the site-saturation mutagenesis described herein, using one or more N,N,G/T cassettes or alternatively using one or more N,N,N cassettes.

The *in vivo* recombination method of the invention can be performed blindly on a pool of unknown hybrids or alleles of a specific polynucleotide or sequence.

However, it is not necessary to know the actual DNA or RNA sequence of the specific polynucleotide.

The approach of using recombination within a mixed population of genes can be useful for the generation of any useful proteins, for example, interleukin I, antibodies, tPA and growth hormone. This approach may be used to generate proteins having altered specificity or activity. The approach may also be useful for the generation of hybrid nucleic acid sequences, for example, promoter regions, introns, exons, enhancer sequences, 31 untranslated regions or 51 untranslated regions of genes. Thus this approach may be used to generate genes having increased rates of expression. This approach may also be useful in the study of repetitive DNA sequences. Finally, this approach may be useful to mutate ribozymes or aptamers.

In one aspect the invention described herein is directed to the use of repeated cycles of reductive reassortment, recombination and selection which allow for the directed molecular evolution of highly complex linear sequences, such as DNA, RNA or proteins thorough recombination.

*Optimized Directed Evolution System*

The invention provides a non-stochastic gene modification system termed "optimized directed evolution system" to generate polypeptides, e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, or antibodies of the invention, with new or altered properties. Optimized directed evolution is directed to the use of repeated cycles of reductive reassortment, recombination and selection that allow for the directed molecular evolution of nucleic acids through recombination. Optimized directed evolution allows generation of a large population of evolved chimeric sequences, wherein the generated population is significantly enriched for sequences that have a predetermined number of crossover events.

A crossover event is a point in a chimeric sequence where a shift in sequence occurs from one parental variant to another parental variant. Such a point is normally at the juncture of where oligonucleotides from two parents are ligated together to form a single sequence. This method allows calculation of the correct concentrations of oligonucleotide sequences so that the final chimeric population of sequences is enriched for the chosen number of crossover events. This provides more control over choosing chimeric variants having a predetermined number of crossover events.

In addition, this method provides a convenient means for exploring a tremendous amount of the possible protein variant space in comparison to other systems.

Previously, if one generated, for example, $10^{13}$ chimeric molecules during a reaction, it would be extremely difficult to test such a high number of chimeric variants for a particular activity. Moreover, a significant portion of the progeny population would have a very high number of crossover events which resulted in proteins that were less likely to have increased levels of a particular activity. By using these methods, the population of chimerics molecules can be enriched for those variants that have a particular number of crossover events. Thus, although one can still generate $10^{13}$ chimeric molecules during a reaction, each of the molecules chosen for further analysis most likely has, for example, only three crossover events. Because the resulting progeny population can be skewed to have a predetermined number of crossover events, the boundaries on the functional variety between the chimeric molecules is reduced. This provides a more manageable number of variables when calculating which oligonucleotide from the original parental polynucleotides might be responsible for affecting a particular trait.

One method for creating a chimeric progeny polynucleotide sequence is to create oligonucleotides corresponding to fragments or portions of each parental sequence. Each oligonucleotide in one aspect includes a unique region of overlap so that mixing the oligonucleotides together results in a new variant that has each oligonucleotide fragment assembled in the correct order. Alternatively protocols for practicing these methods of the invention can be found in U.S. Patent Nos. 6,773,900; 6,740,506; 6,713,282; 6,635,449; 6,605,449; 6,537,776; 6,361,974.

The number of oligonucleotides generated for each parental variant bears a relationship to the total number of resulting crossovers in the chimeric molecule that is ultimately created. For example, three parental nucleotide sequence variants might be provided to undergo a ligation reaction in order to find a chimeric variant having, for example, greater activity at high temperature. As one example, a set of 50 oligonucleotide sequences can be generated corresponding to each portions of each parental variant. Accordingly, during the ligation reassembly process there could be up to 50 crossover events within each of the chimeric sequences. The probability that each of the generated chimeric polynucleotides will contain oligonucleotides from each parental variant in alternating order is very low. If each oligonucleotide fragment is present in the ligation reaction in the same molar quantity it is likely that in some positions oligonucleotides from the same parental polynucleotide will ligate next to one another and thus not result in a crossover event. If the concentration of each oligonucleotide from each parent is kept constant during any ligation step in this example, there is a 1/3 chance

(assuming 3 parents) that an oligonucleotide from the same parental variant will ligate within the chimeric sequence and produce no crossover.

Accordingly, a probability density function (PDF) can be determined to predict the population of crossover events that are likely to occur during each step in a ligation reaction given a set number of parental variants, a number of oligonucleotides corresponding to each variant, and the concentrations of each variant during each step in the ligation reaction. The statistics and mathematics behind determining the PDF is described below. By utilizing these methods, one can calculate such a probability density function, and thus enrich the chimeric progeny population for a predetermined number of crossover events resulting from a particular ligation reaction. Moreover, a target number of crossover events can be predetermined, and the system then programmed to calculate the starting quantities of each parental oligonucleotide during each step in the ligation reaction to result in a probability density function that centers on the predetermined number of crossover events. These methods are directed to the use of repeated cycles of reductive reassortment, recombination and selection that allow for the directed molecular evolution of a nucleic acid encoding a polypeptide through recombination. This system allows generation of a large population of evolved chimeric sequences, wherein the generated population is significantly enriched for sequences that have a predetermined number of crossover events. A crossover event is a point in a chimeric sequence where a shift in sequence occurs from one parental variant to another parental variant. Such a point is normally at the juncture of where oligonucleotides from two parents are ligated together to form a single sequence. The method allows calculation of the correct concentrations of oligonucleotide sequences so that the final chimeric population of sequences is enriched for the chosen number of crossover events. This provides more control over choosing chimeric variants having a predetermined number of crossover events.

In addition, these methods provide a convenient means for exploring a tremendous amount of the possible protein variant space in comparison to other systems. By using the methods described herein, the population of chimerics molecules can be enriched for those variants that have a particular number of crossover events. Thus, although one can still generate $10^{13}$ chimeric molecules during a reaction, each of the molecules chosen for further analysis most likely has, for example, only three crossover events. Because the resulting progeny population can be skewed to have a predetermined number of crossover events, the boundaries on the functional variety between the

chimeric molecules is reduced. This provides a more manageable number of variables when calculating which oligonucleotide from the original parental polynucleotides might be responsible for affecting a particular trait.

In one aspect, the method creates a chimeric progeny polynucleotide sequence by creating oligonucleotides corresponding to fragments or portions of each parental sequence. Each oligonucleotide in one aspect includes a unique region of overlap so that mixing the oligonucleotides together results in a new variant that has each oligonucleotide fragment assembled in the correct order. See also U.S. Patent Nos. 6,537,776; 6,605,449.

*Determining Crossover Events*

Aspects of the invention include a system and software that receive a desired crossover probability density function (PDF), the number of parent genes to be reassembled, and the number of fragments in the reassembly as inputs. The output of this program is a "fragment PDF" that can be used to determine a recipe for producing reassembled genes, and the estimated crossover PDF of those genes. The processing described herein is in one aspect performed in MATLAB™ (The Mathworks, Natick, Massachusetts) a programming language and development environment for technical computing.

*Iterative Processes*

In practicing the invention, these processes can be iteratively repeated. For example, a nucleic acid (or, the nucleic acid) responsible for an altered or new a polypeptide, enzyme, protein, e.g. structural or binding protein, phenotype is identified, re-isolated, again modified, re-tested for activity. This process can be iteratively repeated until a desired phenotype is engineered. For example, an entire biochemical anabolic or catabolic pathway can be engineered into a cell, including, e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, activity.

Similarly, if it is determined that a particular oligonucleotide has no affect at all on the desired trait (e.g., a new a polypeptide, enzyme, protein, e.g. structural or binding protein, phenotype), it can be removed as a variable by synthesizing larger parental oligonucleotides that include the sequence to be removed. Since incorporating the sequence within a larger sequence prevents any crossover events, there will no longer be any variation of this sequence in the progeny polynucleotides. This iterative practice of determining which oligonucleotides are most related to the desired trait, and which are

unrelated, allows more efficient exploration all of the possible protein variants that might be provide a particular trait or activity.

*In vivo shuffling*

*In vivo* shuffling of molecules is use in methods of the invention that provide variants of polypeptides of the invention, e.g., antibodies, a polypeptide, enzyme, protein, e.g. structural or binding protein, and the like. *In vivo* shuffling can be performed utilizing the natural property of cells to recombine multimers. While recombination *in vivo* has provided the major natural route to molecular diversity, genetic recombination remains a relatively complex process that involves 1) the recognition of homologies; 2) strand cleavage, strand invasion, and metabolic steps leading to the production of recombinant chiasma; and finally 3) the resolution of chiasma into discrete recombined molecules. The formation of the chiasma requires the recognition of homologous sequences.

In another aspect, the invention includes a method for producing a hybrid polynucleotide from at least a first polynucleotide and a second polynucleotide. The invention can be used to produce a hybrid polynucleotide by introducing at least a first polynucleotide and a second polynucleotide (e.g., one, or both, being an exemplary polypeptide-, enzyme-, protein-, e.g. structural or binding protein-encoding sequence of the invention) which share at least one region of partial sequence homology into a suitable host cell. The regions of partial sequence homology promote processes which result in sequence reorganization producing a hybrid polynucleotide. The term "hybrid polynucleotide", as used herein, is any nucleotide sequence which results from the method of the present invention and contains sequence from at least two original polynucleotide sequences. Such hybrid polynucleotides can result from intermolecular recombination events which promote sequence integration between DNA molecules. In addition, such hybrid polynucleotides can result from intramolecular reductive reassortment processes which utilize repeated sequences to alter a nucleotide sequence within a DNA molecule.

In *vivo* reassortment is focused on "inter-molecular" processes collectively referred to as "recombination" which in bacteria, is generally viewed as a "RecA-dependent" phenomenon. The invention can rely on recombination processes of a host cell to recombine and re-assort sequences, or the cells' ability to mediate reductive processes to decrease the complexity of quasi-repeated sequences in the cell by deletion.

This process of "reductive reassortment" occurs by an "intra-molecular", RecA-independent process.

Therefore, in another aspect of the invention, novel polynucleotides can be generated by the process of reductive reassortment. The method involves the generation of constructs containing consecutive sequences (original encoding sequences), their insertion into an appropriate vector and their subsequent introduction into an appropriate host cell. The reassortment of the individual molecular identities occurs by combinatorial processes between the consecutive sequences in the construct possessing regions of homology, or between quasi-repeated units. The reassortment process recombines and/or reduces the complexity and extent of the repeated sequences and results in the production of novel molecular species. Various treatments may be applied to enhance the rate of reassortment. These could include treatment with ultra-violet light, or DNA damaging chemicals and/or the use of host cell lines displaying enhanced levels of "genetic instability". Thus the reassortment process may involve homologous recombination or the natural property of quasi-repeated sequences to direct their own evolution.

Repeated or "quasi-repeated" sequences play a role in genetic instability. In the present invention, "quasi-repeats" are repeats that are not restricted to their original unit structure. Quasi-repeated units can be presented as an array of sequences in a construct; consecutive units of similar sequences. Once ligated, the junctions between the consecutive sequences become essentially invisible and the quasi-repetitive nature of the resulting construct is now continuous at the molecular level. The deletion process the cell performs to reduce the complexity of the resulting construct operates between the quasi-repeated sequences. The quasi-repeated units provide a practically limitless repertoire of templates upon which slippage events can occur. The constructs containing the quasi-repeats thus effectively provide sufficient molecular elasticity that deletion (and potentially insertion) events can occur virtually anywhere within the quasi-repetitive units.

When the quasi-repeated sequences are all ligated in the same orientation, for instance head to tail or vice versa, the cell cannot distinguish individual units. Consequently, the reductive process can occur throughout the sequences. In contrast, when for example, the units are presented head to head, rather than head to tail, the inversion delineates the endpoints of the adjacent unit so that deletion formation will favor the loss of discrete units. Thus, it is preferable with the present method that the sequences are in the same orientation. Random orientation of quasi-repeated sequences

will result in the loss of reassortment efficiency, while consistent orientation of the sequences will offer the highest efficiency. However, while having fewer of the contiguous sequences in the same orientation decreases the efficiency, it may still provide sufficient elasticity for the effective recovery of novel molecules. Constructs can be made with the quasi-repeated sequences in the same orientation to allow higher efficiency.

Sequences can be assembled in a head to tail orientation using any of a variety of methods, including the following:

a) Primers that include a poly-A head and poly-T tail which when made single-stranded would provide orientation can be utilized. This is accomplished by having the first few bases of the primers made from RNA and hence easily removed RNaseH.

b) Primers that include unique restriction cleavage sites can be utilized. Multiple sites, a battery of unique sequences and repeated synthesis and ligation steps would be required.

c) The inner few bases of the primer could be thiolated and an exonuclease used to produce properly tailed molecules.

The recovery of the re-assorted sequences relies on the identification of cloning vectors with a reduced repetitive index (RI). The re-assorted encoding sequences can then be recovered by amplification. The products are re-cloned and expressed. The recovery of cloning vectors with reduced RI can be affected by:

1) The use of vectors only stably maintained when the construct is reduced in complexity.

2) The physical recovery of shortened vectors by physical procedures. In this case, the cloning vector would be recovered using standard plasmid isolation procedures and size fractionated on either an agarose gel, or column with a low molecular weight cut off utilizing standard procedures.

3) The recovery of vectors containing interrupted genes which can be selected when insert size decreases.

4) The use of direct selection techniques with an expression vector and the appropriate selection.

Encoding sequences (for example, genes) from related organisms may demonstrate a high degree of homology and encode quite diverse protein products. These types of sequences are particularly useful in the present invention as quasi-repeats.

However, while the examples illustrated below demonstrate the reassortment of nearly identical original encoding sequences (quasi-repeats), this process is not limited to such nearly identical repeats.

The following example demonstrates a method of the invention. Encoding nucleic acid sequences (quasi-repeats) derived from three (3) unique species are described. Each sequence encodes a protein with a distinct set of properties. Each of the sequences differs by a single or a few base pairs at a unique position in the sequence. The quasi-repeated sequences are separately or collectively amplified and ligated into random assemblies such that all possible permutations and combinations are available in the population of ligated molecules. The number of quasi-repeat units can be controlled by the assembly conditions. The average number of quasi-repeated units in a construct is defined as the repetitive index (RI).

Once formed, the constructs may, or may not be size fractionated on an agarose gel according to published protocols, inserted into a cloning vector and transfected into an appropriate host cell. The cells are then propagated and "reductive reassortment" is effected. The rate of the reductive reassortment process may be stimulated by the introduction of DNA damage if desired. Whether the reduction in RI is mediated by deletion formation between repeated sequences by an "intra-molecular" mechanism, or mediated by recombination-like events through "inter-molecular" mechanisms is immaterial. The end result is a reassortment of the molecules into all possible combinations.

Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual shuffled library members having the ability to bind or otherwise interact, or catalyze a particular reaction (*e.g.*, such as catalytic domain of an enzyme) with a predetermined macromolecule, such as for example a proteinaceous receptor, an oligosaccharide, virion, or other predetermined compound or structure.

The polypeptides that are identified from such libraries can be used for therapeutic, diagnostic, research and related purposes (*e.g.*, catalysts, solutes for increasing osmolarity of an aqueous solution and the like) and/or can be subjected to one or more additional cycles of shuffling and/or selection.

In another aspect, it is envisioned that prior to or during recombination or reassortment, polynucleotides generated by the method of the invention can be subjected to agents or processes which promote the introduction of mutations into the original

polynucleotides. The introduction of such mutations would increase the diversity of resulting hybrid polynucleotides and polypeptides encoded therefrom. The agents or processes which promote mutagenesis can include, but are not limited to: (+)-CC-1065, or a synthetic analog such as (+)-CC-1065-(N3-Adenine (*See* Sun and Hurley, (1992); an N-acetylated or deacetylated 4'-fluro-4-aminobiphenyl adduct capable of inhibiting DNA synthesis (*See* , for example, van de Poll *et al.* (1992)); or a N-acetylated or deacetylated 4-aminobiphenyl adduct capable of inhibiting DNA synthesis (*See* also, van de Poll *et al.* (1992), pp. 751-758); trivalent chromium, a trivalent chromium salt, a polycyclic aromatic hydrocarbon (PAH) DNA adduct capable of inhibiting DNA replication, such as 7-bromomethyl-benz[*a*]anthracene ("BMA"), tris(2,3-dibromopropyl)phosphate ("Tris-BP"), 1,2-dibromo-3-chloropropane ("DBCP"), 2-bromoacrolein (2BA), benzo[*a*]pyrene-7,8-dihydrodiol-9-10-epoxide ("BPDE"), a platinum(II) halogen salt, N-hydroxy-2-amino-3-methylimidazo[4,5-*f*]-quinoline ("N-hydroxy-IQ") and N-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-*f*]-pyridine ("N-hydroxy-PhIP"). Exemplary means for slowing or halting PCR amplification consist of UV light (+)-CC-1065 and (+)-CC-1065-(N3-Adenine). Particularly encompassed means are DNA adducts or polynucleotides comprising the DNA adducts from the polynucleotides or polynucleotides pool, which can be released or removed by a process including heating the solution comprising the polynucleotides prior to further processing.

In another aspect the invention is directed to a method of producing recombinant proteins having biological activity by treating a sample comprising double-stranded template polynucleotides encoding a wild-type protein under conditions according to the invention which provide for the production of hybrid or re-assorted polynucleotides.

*Producing sequence variants*

The invention also provides additional methods for making sequence variants of the nucleic acid (e.g., polypeptide, enzyme, protein, e.g. structural or binding protein) sequences of the invention. The invention also provides additional methods for isolating a polypeptide, enzyme, protein, e.g. structural or binding protein, using the nucleic acids and polypeptides of the invention. In one aspect, the invention provides for variants of a polypeptide, enzyme, protein, e.g. structural or binding protein, coding sequence (e.g., a gene, cDNA or message) of the invention, which can be altered by any means, including, e.g., random or stochastic methods, or, non-stochastic, or "directed evolution," methods, as described above.

The isolated variants may be naturally occurring. Variant can also be created *in vitro*. Variants may be created using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives may be created using chemical synthesis or modification procedures. Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids which encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. These nucleotide differences can result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants may be created using error prone PCR. In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Error prone PCR is described, e.g., in Leung (1989) Technique 1:11-15) and Caldwell (1992) PCR Methods Applic. 2:28-33. Briefly, in such procedures, nucleic acids to be mutagenized are mixed with PCR primers, reaction buffer, $MgCl_2$, $MnCl_2$, Taq polymerase and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction may be performed using 20 fmoles of nucleic acid to be mutagenized, 30 pmole of each PCR primer, a reaction buffer comprising 50mM KCl, 10mM Tris HCl (pH 8.3) and 0.01% gelatin, 7mM MgCl2, 0.5mM $MnCl_2$, 5 units of Taq polymerase, 0.2mM dGTP, 0.2mM dATP, 1mM dCTP, and 1mM dTTP. PCR may be performed for 30 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min. However, it will be appreciated that these parameters may be varied as appropriate. The mutagenized nucleic acids are cloned into an appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids are evaluated.

Variants may also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described, e.g., in Reidhaar-Olson (1988) Science 241:53-57. Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the

cloned DNA to be mutagenized. Clones containing the mutagenized DNA are recovered and the activities of the polypeptides they encode are assessed.

Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in, e.g., U.S. Patent No. 5,965,408.

Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different but highly related DNA sequence *in vitro*, as a result of random fragmentation of the DNA molecule based on sequence homology, followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described, e.g., in Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751. Briefly, in such procedures a plurality of nucleic acids to be recombined are digested with DNase to generate fragments having an average size of 50-200 nucleotides. Fragments of the desired average size are purified and resuspended in a PCR mixture. PCR is conducted under conditions which facilitate recombination between the nucleic acid fragments. For example, PCR may be performed by resuspending the purified fragments at a concentration of 10-30ng/μl in a solution of 0.2mM of each dNTP, 2.2mM $MgCl_2$, 50mM KCL, 10mM Tris HCl, pH 9.0, and 0.1% Triton X-100. 2.5 units of Taq polymerase per 100:l of reaction mixture is added and PCR is performed using the following regime: 94°C for 60 seconds, 94°C for 30 seconds, 50-55°C for 30 seconds, 72°C for 30 seconds (30-45 times) and 72°C for 5 minutes. However, it will be appreciated that these parameters may be varied as appropriate. In some aspects, oligonucleotides may be included in the PCR reactions. In other aspects, the Klenow fragment of DNA polymerase I may be used in a first set of PCR reactions and Taq polymerase may be used in a subsequent set of PCR reactions. Recombinant sequences are isolated and the activities of the polypeptides they encode are assessed.

Variants may also be created by *in vivo* mutagenesis. In some aspects, random mutations in a sequence of interest are generated by propagating the sequence of interest in a bacterial strain, such as an E. coli strain, which carries mutations in one or more of the DNA repair pathways. Such "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for

use for *in vivo* mutagenesis are described in PCT Publication No. WO 91/16427, published October 31, 1991, entitled "Methods for Phenotype Creation from Multiple Gene Populations".

Variants may also be generated using cassette mutagenesis. In cassette mutagenesis a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

Recursive ensemble mutagenesis may also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described, e.g., in Arkin (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815.

In some aspects, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described, e.g., in Delegrave (1993) Biotechnology Res. 11:1548-1552. Random and site-directed mutagenesis are described, e.g., in Arnold (1993) Current Opinion in Biotechnology 4:450-455.

In some aspects, the variants are created using shuffling procedures wherein portions of a plurality of nucleic acids which encode distinct polypeptides are fused together to create chimeric nucleic acid sequences which encode chimeric polypeptides as described in U.S. Patent No. 5,965,408, filed July 9, 1996, entitled, "Method of DNA Reassembly by Interrupting Synthesis" and U.S. Patent No. 5,939,250, filed May 22, 1996, entitled, "Production of Enzymes Having Desired Activities by Mutagenesis.

The variants of the polypeptides of the invention may be variants in which one or more of the amino acid residues of the polypeptides of the sequences of the invention are substituted with a conserved or non-conserved amino acid residue (in one aspect a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.

Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the following replacements: replacements of an aliphatic amino acid such as Alanine, Valine, Leucine and Isoleucine with another aliphatic amino acid; replacement of a Serine with a Threonine or vice versa; replacement of an acidic residue such as Aspartic acid and Glutamic acid with another acidic residue; replacement of a residue bearing an amide group, such as Asparagine and Glutamine, with another residue bearing an amide group; exchange of a basic residue such as Lysine and Arginine with another basic residue; and replacement of an aromatic residue such as Phenylalanine, Tyrosine with another aromatic residue.

Other variants are those in which one or more of the amino acid residues of a polypeptide of the invention includes a substituent group.

Still other variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Additional variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretory sequence, a proprotein sequence or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

In some aspects, the fragments, derivatives and analogs retain the same biological function or activity as the polypeptides of the invention. In other aspects, the fragment, derivative, or analog includes a proprotein, such that the fragment, derivative, or analog can be activated by cleavage of the proprotein portion to produce an active polypeptide.

*Optimizing codons to achieve high levels of protein expression in host cells*

The invention provides methods for modifying polypeptide-, enzyme-, protein-, e.g. structural or binding protein-encoding nucleic acids to modify codon usage. In one aspect, the invention provides methods for modifying codons in a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, to increase or decrease its expression in a host cell. The invention also provides nucleic acids encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, modified to increase its expression in a host cell, a polypeptide, enzyme, protein, e.g. structural or binding protein, so modified, and methods of making the modified a polypeptide, enzyme, protein, e.g. structural or binding protein. The method comprises identifying a "non-preferred" or a "less preferred" codon in a polypeptide-, enzyme-, protein-, e.g. structural

or binding protein-encoding nucleic acid and replacing one or more of these non-preferred or less preferred codons with a "preferred codon" encoding the same amino acid as the replaced codon and at least one non- preferred or less preferred codon in the nucleic acid has been replaced by a preferred codon encoding the same amino acid. A preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non- preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell.

Host cells for expressing the nucleic acids, expression cassettes and vectors of the invention include bacteria, yeast, fungi, plant cells, insect cells and mammalian cells. Thus, the invention provides methods for optimizing codon usage in all of these cells, codon-altered nucleic acids and polypeptides made by the codon-altered nucleic acids. Exemplary host cells include gram negative bacteria, such as *Escherichia coli*; gram positive bacteria, such as *Streptomyces* sp., *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, *Bacillus subtilis, Bacillus cereus*. Exemplary host cells also include eukaryotic organisms, e.g., various yeast, such as *Saccharomyces* sp., including *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, and *Kluyveromyces lactis*, *Hansenula polymorpha*, *Aspergillus niger*, and mammalian cells and cell lines and insect cells and cell lines. Thus, the invention also includes nucleic acids and polypeptides optimized for expression in these organisms and species.

For example, the codons of a nucleic acid encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, isolated from a bacterial cell are modified such that the nucleic acid is optimally expressed in a bacterial cell different from the bacteria from which the polypeptide, enzyme, protein, e.g. structural or binding protein was derived, a yeast, a fungi, a plant cell, an insect cell or a mammalian cell. Methods for optimizing codons are well known in the art, see, e.g., U.S. Patent No. 5,795,737; Baca (2000) Int. J. Parasitol. 30:113-118; Hale (1998) Protein Expr. Purif. 12:185-188; Narum (2001) Infect. Immun. 69:7250-7253. See also Narum (2001) Infect. Immun. 69:7250-7253, describing optimizing codons in mouse systems; Outchkourov (2002) Protein Expr. Purif. 24:18-24, describing optimizing codons in yeast; Feng (2000) Biochemistry 39:15399-15409, describing optimizing codons in *E. coli*; Humphreys (2000) Protein Expr. Purif. 20:252-264, describing optimizing codon usage that affects secretion in *E. coli*.

Transgenic non-human animals

The invention provides transgenic non-human animals comprising a nucleic acid, a polypeptide (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein,), an expression cassette or vector or a transfected or transformed cell of the invention. The invention also provides methods of making and using these transgenic non-human animals.

The transgenic non-human animals can be, e.g., goats, rabbits, sheep, pigs (including all swine, hogs and related animals), cows, rats and mice, comprising the nucleic acids of the invention. These animals can be used, e.g., as *in vivo* models to study a polypeptide, enzyme, protein, e.g. structural or binding protein, activity, or, as models to screen for agents that change the polypeptide, enzyme, protein, e.g. structural or binding protein activity *in vivo*. The coding sequences for the polypeptides to be expressed in the transgenic non-human animals can be designed to be constitutive, or, under the control of tissue-specific, developmental-specific or inducible transcriptional regulatory factors. Transgenic non-human animals can be designed and generated using any method known in the art; see, e.g., U.S. Patent Nos. 6,211,428; 6,187,992; 6,156,952; 6,118,044; 6,111,166; 6,107,541; 5,959,171; 5,922,854; 5,892,070; 5,880,327; 5,891,698; 5,639,940; 5,573,933; 5,387,742; 5,087,571, describing making and using transformed cells and eggs and transgenic mice, rats, rabbits, sheep, pigs and cows. See also, e.g., Pollock (1999) J. Immunol. Methods 231:147-157, describing the production of recombinant proteins in the milk of transgenic dairy animals; Baguisi (1999) Nat. Biotechnol. 17:456-461, demonstrating the production of transgenic goats. U.S. Patent No. 6,211,428, describes making and using transgenic non-human mammals which express in their brains a nucleic acid construct comprising a DNA sequence. U.S. Patent No. 5,387,742, describes injecting cloned recombinant or synthetic DNA sequences into fertilized mouse eggs, implanting the injected eggs in pseudo-pregnant females, and growing to term transgenic mice. U.S. Patent No. 6,187,992, describes making and using a transgenic mouse.

"Knockout animals" can also be used to practice the methods of the invention. For example, in one aspect, the transgenic or modified animals of the invention comprise a "knockout animal," e.g., a "knockout mouse," engineered not to express an endogenous gene, which is replaced with a gene expressing a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention, or, a fusion protein

comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention.

Transgenic Plants and Seeds

The invention provides transgenic plants and seeds comprising a nucleic acid, a polypeptide (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein,), an expression cassette or vector or a transfected or transformed cell of the invention. The invention also provides plant products, e.g., oils, seeds, leaves, extracts and the like, comprising a nucleic acid and/or a polypeptide (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein,) of the invention. The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot). The invention also provides methods of making and using these transgenic plants and seeds. The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with any method known in the art. See, for example, U.S. Patent No. 6,309,872.

Nucleic acids and expression constructs of the invention can be introduced into a plant cell by any means. For example, nucleic acids or expression constructs can be introduced into the genome of a desired plant host, or, the nucleic acids or expression constructs can be episomes. Introduction into the genome of a desired plant can be such that the host's a polypeptide, enzyme, protein, e.g. structural or binding protein, production is regulated by endogenous transcriptional or translational control elements. The invention also provides "knockout plants" where insertion of gene sequence by, e.g., homologous recombination, has disrupted the expression of the endogenous gene. Means to generate "knockout" plants are well-known in the art, see, e.g., Strepp (1998) Proc Natl. Acad. Sci. USA 95:4368-4373; Miao (1995) Plant J 7:359-365. See discussion on transgenic plants, below.

The nucleic acids of the invention can be used to confer desired traits on essentially any plant, e.g., on starch-producing plants, such as potato, wheat, rice, barley, and the like. Nucleic acids of the invention can be used to manipulate metabolic pathways of a plant in order to optimize or alter host's expression of polypeptide, enzyme, protein, e.g. structural or binding protein. The can change a polypeptide, enzyme, protein, e.g. structural or binding protein, activity in a plant. Alternatively, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can be used in production of a transgenic plant to produce a compound not naturally produced by that plant. This can lower production costs or create a novel product.

In one aspect, the first step in production of a transgenic plant involves making an expression construct for expression in a plant cell. These techniques are well known in the art. They can include selecting and cloning a promoter, a coding sequence for facilitating efficient binding of ribosomes to mRNA and selecting the appropriate gene terminator sequences. One exemplary constitutive promoter is CaMV35S, from the cauliflower mosaic virus, which generally results in a high degree of expression in plants. Other promoters are more specific and respond to cues in the plant's internal or external environment. An exemplary light-inducible promoter is the promoter from the cab gene, encoding the major chlorophyll a/b binding protein.

In one aspect, the nucleic acid is modified to achieve greater expression in a plant cell. For example, a sequence of the invention is likely to have a higher percentage of A-T nucleotide pairs compared to that seen in a plant, some of which prefer G-C nucleotide pairs. Therefore, A-T nucleotides in the coding sequence can be substituted with G-C nucleotides without significantly changing the amino acid sequence to enhance production of the gene product in plant cells.

Selectable marker gene can be added to the gene construct in order to identify plant cells or tissues that have successfully integrated the transgene. This may be necessary because achieving incorporation and expression of genes in plant cells is a rare event, occurring in just a few percent of the targeted tissues or cells. Selectable marker genes encode proteins that provide resistance to agents that are normally toxic to plants, such as antibiotics or herbicides. Only plant cells that have integrated the selectable marker gene will survive when grown on a medium containing the appropriate antibiotic or herbicide. As for other inserted genes, marker genes also require promoter and termination sequences for proper function.

In one aspect, making transgenic plants or seeds comprises incorporating sequences of the invention and, optionally, marker genes into a target expression construct (e.g., a plasmid), along with positioning of the promoter and the terminator sequences. This can involve transferring the modified gene into the plant through a suitable method. For example, a construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. For example, see, e.g., Christou (1997) Plant Mol. Biol. 35:197-203; Pawlowski (1996) Mol. Biotechnol. 6:17-30; Klein (1987) Nature 327:70-73; Takumi (1997) Genes Genet. Syst. 72:63-69, discussing use of

particle bombardment to introduce transgenes into wheat; and Adam (1997) supra, for use of particle bombardment to introduce YACs into plant cells. For example, Rinehart (1997) supra, used particle bombardment to generate transgenic cotton plants. Apparatus for accelerating particles is described U.S. Pat. No. 5,015,580; and, the commercially available BioRad (Biolistics) PDS-2000 particle acceleration instrument; see also, John, U.S. Patent No. 5,608,148; and Ellis, U.S. Patent No. 5, 681,730, describing particle-mediated transformation of gymnosperms.

In one aspect, protoplasts can be immobilized and injected with a nucleic acids, e.g., an expression construct. Although plant regeneration from protoplasts is not easy with cereals, plant regeneration is possible in legumes using somatic embryogenesis from protoplast derived callus. Organized tissues can be transformed with naked DNA using gene gun technique, where DNA is coated on tungsten microprojectiles, shot 1/100th the size of cells, which carry the DNA deep into cells and organelles. Transformed tissue is then induced to regenerate, usually by somatic embryogenesis. This technique has been successful in several cereal species including maize and rice.

Nucleic acids, e.g., expression constructs, can also be introduced in to plant cells using recombinant viruses. Plant cells can be transformed using viral vectors, such as, e.g., tobacco mosaic virus derived vectors (Rouwendal (1997) Plant Mol. Biol. 33:989-999), see Porta (1996) "Use of viral replicons for the expression of genes in plants," Mol. Biotechnol. 5:209-221.

Alternatively, nucleic acids, e.g., an expression construct, can be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, e.g., Horsch (1984) *Science* 233:496-498; Fraley (1983) *Proc. Natl. Acad. Sci. USA* 80:4803 (1983); *Gene Transfer to Plants*, Potrykus, ed. (Springer-Verlag, Berlin 1995). The DNA in an A. *tumefaciens* cell is contained in the bacterial chromosome as well as in another structure known as a Ti (tumor-inducing) plasmid. The Ti plasmid contains a stretch of DNA termed T-DNA (~20 kb long) that is transferred to the plant cell in the infection process and a series of vir (virulence) genes that direct the infection process. *A. tumefaciens* can only infect a plant through wounds: when a plant root or stem is wounded it gives off certain chemical

signals, in response to which, the vir genes of *A. tumefaciens* become activated and direct a series of events necessary for the transfer of the T-DNA from the Ti plasmid to the plant's chromosome. The T-DNA enters the plant cell through the wound. One speculation is that the T-DNA waits until the plant DNA is being replicated or transcribed, then inserts itself into the exposed plant DNA. In order to use *A. tumefaciens* as a transgene vector, the tumor-inducing section of T-DNA have to be removed, while retaining the T-DNA border regions and the vir genes. The transgene is then inserted between the T-DNA border regions, where it is transferred to the plant cell and becomes integrated into the plant's chromosomes.

The invention provides for the transformation of monocotyledonous plants using the nucleic acids of the invention, including important cereals, see Hiei (1997) Plant Mol. Biol. 35:205-218. See also, *e.g.*, Horsch, Science (1984) 233:496; Fraley (1983) Proc. Natl. Acad. Sci USA 80:4803; Thykjaer (1997) supra; Park (1996) Plant Mol. Biol. 32:1135-1148, discussing T-DNA integration into genomic DNA. See also D'Halluin, U.S. Patent No. 5,712,135, describing a process for the stable integration of a DNA comprising a gene that is functional in a cell of a cereal, or other monocotyledonous plant.

In one aspect, the third step can involve selection and regeneration of whole plants capable of transmitting the incorporated target gene to the next generation. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker that has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture*, *Handbook of Plant Cell Culture*, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants*, *Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee (1987) Ann. Rev. of Plant Phys. 38:467-486. To obtain whole plants from transgenic tissues such as immature embryos, they can be grown under controlled environmental conditions in a series of media containing nutrients and hormones, a process known as tissue culture. Once whole plants are generated and produce seed, evaluation of the progeny begins.

After the expression cassette is stably incorporated in transgenic plants, it can be introduced into other plants by sexual crossing. Any of a number of standard

breeding techniques can be used, depending upon the species to be crossed. Since transgenic expression of the nucleic acids of the invention leads to phenotypic changes, plants comprising the recombinant nucleic acids of the invention can be sexually crossed with a second plant to obtain a final product. Thus, the seed of the invention can be derived from a cross between two transgenic plants of the invention, or a cross between a plant of the invention and another plant. The desired effects (e.g., expression of the polypeptides of the invention to produce a plant in which flowering behavior is altered) can be enhanced when both parental plants express the polypeptides (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein,) of the invention. The desired effects can be passed to future plant generations by standard propagation means.

The nucleic acids and polypeptides of the invention are expressed in or inserted in any plant or seed. Transgenic plants of the invention can be dicotyledonous or monocotyledonous. Examples of monocot transgenic plants of the invention are grasses, such as meadow grass (blue grass, *Poa*), forage grass such as festuca, lolium, temperate grass, such as *Agrostis*, and cereals, e.g., wheat, oats, rye, barley, rice, sorghum, and maize (corn). Examples of dicot transgenic plants of the invention are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family *Brassicaceae*), such as cauliflower, rape seed, and the closely related model organism *Arabidopsis thaliana*. Thus, the transgenic plants and seeds of the invention include a broad range of plants, including, but not limited to, species from the genera *Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lupinus, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannisetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna*, and *Zea*.

In alternative embodiments, the nucleic acids of the invention are expressed in plants which contain fiber cells, including, *e.g.*, cotton, silk cotton tree (Kapok, Ceiba pentandra), desert willow, creosote bush, winterfat, balsa, ramie, kenaf, hemp, roselle, jute, sisal abaca and flax. In alternative embodiments, the transgenic plants of the invention can be members of the genus *Gossypium*, including members of any *Gossypium* species, such as *G. arboreum*;. *G. herbaceum*, *G. barbadense*, and *G. hirsutum*.

The invention also provides for transgenic plants to be used for producing large amounts of the polypeptides (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, or antibody) of the invention. For example, see Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296 (producing human milk protein beta-casein in transgenic potato plants using an auxin-inducible, bidirectional mannopine synthase (mas1',2') promoter with *Agrobacterium tumefaciens*-mediated leaf disc transformation methods).

Using known procedures, one of skill can screen for plants of the invention by detecting the increase or decrease of transgene mRNA or protein in transgenic plants. Means for detecting and quantitation of mRNAs or proteins are well known in the art.

Polypeptides and peptides

In one aspect, the invention provides isolated, synthetic or recombinant polypeptides having a sequence identity (e.g., at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity, or homology) to an exemplary sequence of the invention, e.g., proteins having the sequence of (a sequence as set forth in) SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, etc., and all polypeptides (amino acid sequences) disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699). The percent sequence identity can be over the full length of the polypeptide, or, the identity can be over a region of at least about 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or more residues.

"Amino acid" or "amino acid sequence" as used herein refer to an oligopeptide, peptide, polypeptide, or protein sequence, or to a fragment, portion, or subunit of any of these and to naturally occurring or synthetic molecules. "Amino acid" or "amino acid sequence" include an oligopeptide, peptide, polypeptide, or protein sequence, or to a fragment, portion, or subunit of any of these, and to naturally occurring or synthetic molecules. The term "polypeptide" as used herein, refers to amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres and may contain modified amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Modifications can

occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide may have many types of modifications. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pegylation, glucan hydrolase processing, phosphorylation, prenylation, racemization, selenoylation, sulfation and transfer-RNA mediated addition of amino acids to protein such as arginylation. (*See* Creighton, T.E., Proteins – Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)). The peptides and polypeptides of the invention also include all "mimetic" and "peptidomimetic" forms, as described in further detail, below.

As used herein, the term "isolated" means that the material is removed from its original environment (*e.g.*, the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition and still be isolated in that such vector or composition is not part of its natural environment. As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual nucleic acids obtained from a library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. In one embodiment, purified nucleic acids of the invention have been purified from the remainder of the genomic DNA in the organism by at least $10^4$-$10^6$ fold. However, the term "purified" also includes nucleic acids which have been purified from the remainder of the genomic DNA or from other sequences in a library or other environment by at least one

order of magnitude, typically two or three orders and more typically four or five orders of magnitude.

"Recombinant" polypeptides or proteins refer to polypeptides or proteins produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide or protein. "Synthetic" polypeptides or proteins of the invention can be those prepared by chemical synthesis. Solid-phase chemical peptide synthesis methods can also be used to synthesize the polypeptide or fragments of the invention. Such method have been known in the art since the early 1960's (Merrifield, R. B., *J. Am. Chem. Soc.*, 85:2149-2154, 1963) (See also Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) and have recently been employed in commercially available laboratory peptide design and synthesis kits (Cambridge Research Biochemicals). Such commercially available laboratory kits have generally utilized the teachings of H. M. Geysen *et al, Proc. Natl. Acad. Sci., USA*, 81:3998 (1984) and provide for synthesizing peptides upon the tips of a multitude of "rods" or "pins" all of which are connected to a single plate.

The invention provides polypeptides and peptides having amino acid sequences that are "substantially identical" to a polypeptide of the invention, e.g., an exemplary sequence of the invention; in one aspect, a "substantially identical" sequence comprises an amino acid sequence that differs from a reference sequence (exemplary sequence of the invention) by one or more conservative or non-conservative amino acid substitutions, deletions, or insertions. In one aspect, the substitution occurs at a site that is not the active site of the molecule, or, alternatively the substitution occurs at a site that is the active site of the molecule, provided that the polypeptide essentially retains its functional (enzymatic) properties. A conservative amino acid substitution, for example, substitutes one amino acid for another of the same class (*e.g.*, substitution of one hydrophobic amino acid, such as isoleucine, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid or glutamine for asparagine). One or more amino acids can be deleted, for example, from a polypeptide, resulting in modification of the structure of the polypeptide, without significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for a polypeptide, enzyme, protein, e.g. structural or binding protein, biological activity can be removed. Modified polypeptide sequences of the invention can be assayed for enzyme, structural or binding activity by any number of methods, including contacting the modified polypeptide sequence with a

substrate and determining whether the modified polypeptide decreases the amount of specific substrate in the assay or increases the bioproducts of the reaction of a functional polypeptide, enzyme, protein, e.g. structural or binding protein, with the substrate. Assays for enzyme activity, binding activity and/or structural activity are well known in the art.

"Fragments" as used herein are a portion of a naturally occurring protein which can exist in at least two different conformations. Fragments can have the same or substantially the same amino acid sequence as the naturally occurring protein. Fragments which have different three dimensional structures as the naturally occurring protein are also included. An example of this, is a "pro-form" molecule, such as a low activity proprotein that can be modified by cleavage to produce a mature enzyme with significantly higher activity.

Polypeptides or peptides of the invention can also be shorter than the full length of exemplary polypeptides. In alternative aspects, the invention provides polypeptides (peptides, fragments) ranging in size between about 5 and the full length of a polypeptide, e.g., an enzyme, such as a polypeptide, enzyme, protein, e.g. structural or binding protein,; exemplary sizes being of about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, or more residues, e.g., contiguous residues of an exemplary a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention. Peptides of the invention (e.g., a subsequence of an exemplary polypeptide of the invention) can be useful as, e.g., labeling probes, antigens, toleragens, motifs, a polypeptide, enzyme, protein, e.g. structural or binding protein, active sites (e.g., "catalytic domains"), signal sequences and/or prepro domains.

In alternative aspects, polypeptides of the invention having enzyme, structural or binding activity are members of a genus of polypeptides sharing specific structural elements, e.g., amino acid residues, that correlate with enzyme, structural or binding activity. These shared structural elements can be used for the routine generation of polypeptide, enzyme, protein, e.g. structural or binding protein, variants. These shared structural elements of a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can be used as guidance for the routine generation of a polypeptide, enzyme, protein, e.g. structural or binding protein, variants within the scope of the genus of polypeptides of the invention.

Polypeptides and peptides of the invention can be isolated from natural sources, be synthetic, or be recombinantly generated polypeptides. Peptides and proteins can be recombinantly expressed *in vitro* or *in vivo*. The peptides and polypeptides of the invention can be made and isolated using any method known in the art. Polypeptide and peptides of the invention can also be synthesized, whole or in part, using chemical methods well known in the art. See e.g., Caruthers (1980) Nucleic Acids Res. Symp. Ser. 215-223; Horn (1980) Nucleic Acids Res. Symp. Ser. 225-232; Banga, A.K., Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems (1995) Technomic Publishing Co., Lancaster, PA. For example, peptide synthesis can be performed using various solid-phase techniques (see e.g., Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) and automated synthesis may be achieved, e.g., using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

In some embodiments, the isolated, synthetic or recombinant polypeptide of the invention have an enzyme activity, a binding activity and/or a structural activity which is thermostable or thermotolerant, or the isolated, synthetic or recombinant nucleic acids encode a polypeptide having an enzyme activity, a binding activity and/or a structural activity which is thermostable or thermotolerant. In alternative embodiments, a thermostable polypeptide of the invention can retain an enzyme activity, a binding activity and/or a structural activity, under conditions comprising a temperature range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 37°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. The thermostable polypeptides according to the invention can retain an enzyme activity, a binding activity and/or a structural activity in temperatures in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C,

about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. In some embodiments, the thermostable polypeptides of this invention retains an enzyme activity, a binding activity and/or a structural activity at a temperature in the ranges described above, at about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about pH 8.5, about pH 9.0, about pH 9.5, about pH 10.0, about pH 10.5, about pH 11.0, about pH 11.5, about pH 12.0 or more.

In other embodiments, the isolated, synthetic or recombinant polypeptide of the invention has an enzyme activity, a binding activity and/or a structural activity which is thermotolerant, or the isolated, synthetic or recombinant nucleic acids encode a polypeptide having an enzyme activity, a binding activity and/or a structural activity which is thermotolerant. The thermotolerant polypeptides of the invention can retain an enzyme activity, a binding activity and/or a structural activity after exposure to conditions comprising a temperature in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more. The thermotolerant polypeptides of the invention can retain an enzyme activity, a binding activity and/or a structural activity after exposure to a temperature in the range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C

or more. In some embodiments, the thermotolerant polypeptides of the invention retain an enzyme activity, a binding activity and/or a structural activity after exposure to a temperature in the ranges described above, at about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about pH 8.5, about pH 9.0, about pH 9.5, about pH 10.0, about pH 10.5, about pH 11.0, about pH 11.5, about pH 12.0 or more.

In some embodiments, the polypeptide can retain an enzyme activity, a binding activity and/or a structural activity under conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 or less (more acidic) pH. In other embodiments, the polypeptide can retain an enzyme activity, a binding activity and/or a structural activity under conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11.0, pH 11.5, pH 12, pH 12.5 or more (more basic) pH. In some embodiments, the polypeptide can retain an enzyme activity, a binding activity and/or a structural activity after exposure to conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 or less (more acidic) pH. In other embodiments, the polypeptide can retain an enzyme activity, a binding activity and/or a structural activity after exposure to conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11.0, pH 11.5, pH 12, pH 12.5 or more (more basic) pH.

The peptides and polypeptides of the invention can also be glycosylated. The glycosylation can be added post-translationally either chemically or by cellular biosynthetic mechanisms, wherein the later incorporates the use of known glycosylation motifs, which can be native to the sequence or can be added as a peptide or added in the nucleic acid coding sequence. The glycosylation can be O-linked or N-linked.

The peptides and polypeptides of the invention, as defined above, include all "mimetic" and "peptidomimetic" forms. The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound which has substantially the same structural and/or functional characteristics of the polypeptides of the invention. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. As with polypeptides of the invention which are conservative variants or members of a genus of polypeptides of the invention (e.g., having about 50%

or more sequence identity to an exemplary sequence of the invention), routine experimentation will determine whether a mimetic is within the scope of the invention, i.e., that its structure and/or function is not substantially altered. Thus, in one aspect, a mimetic composition is within the scope of the invention if it has a polypeptide, enzyme, protein, e.g. structural or binding protein's activity.

Polypeptide mimetic compositions of the invention can contain any combination of non-natural structural components. In alternative aspect, mimetic compositions of the invention include one or all of the following three structural groups: a) residue linkage groups other than the natural amide bond ("peptide bond") linkages; b) non-natural residues in place of naturally occurring amino acid residues; or c) residues which induce secondary structural mimicry, i.e., to induce or stabilize a secondary structure, e.g., a beta turn, gamma turn, beta sheet, alpha helix conformation, and the like. For example, a polypeptide of the invention can be characterized as a mimetic when all or some of its residues are joined by chemical means other than natural peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as, e.g., glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-diisopropylcarbodiimide (DIC). Linking groups that can be an alternative to the traditional amide bond ("peptide bond") linkages include, e.g., ketomethylene (e.g., -C(=O)-$CH_2$- for -C(=O)-NH-), aminomethylene ($CH_2$-NH), ethylene, olefin (CH=CH), ether ($CH_2$-O), thioether ($CH_2$-S), tetrazole ($CN_4$-), thiazole, retroamide, thioamide, or ester (see, e.g., Spatola (1983) in Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY).

A polypeptide of the invention can also be characterized as a mimetic by containing all or some non-natural residues in place of naturally occurring amino acid residues. Non-natural residues are well described in the scientific and patent literature; a few exemplary non-natural compositions useful as mimetics of natural amino acid residues and guidelines are described below. Mimetics of aromatic amino acids can be generated by replacing by, e.g., D- or L- naphylalanine; D- or L- phenylglycine; D- or L-2 thieneylalanine; D- or L-1, -2, 3-, or 4- pyreneylalanine; D- or L-3 thieneylalanine; D- or L-(2-pyridinyl)-alanine; D- or L-(3-pyridinyl)-alanine; D- or L-(2-pyrazinyl)-alanine; D- or L-(4-isopropyl)-phenylglycine; D-(trifluoromethyl)-phenylglycine; D-(trifluoromethyl)-phenylalanine; D-p-fluoro-phenylalanine; D- or L-p-biphenylphenylalanine; D- or L-p-methoxy-biphenylphenylalanine; D- or L-2-

indole(alkyl)alanines; and, D- or L-alkylainines, where alkyl can be substituted or unsubstituted methyl, ethyl, propyl, hexyl, butyl, pentyl, isopropyl, iso-butyl, sec-isotyl, iso-pentyl, or a non-acidic amino acids. Aromatic rings of a non-natural amino acid include, e.g., thiazolyl, thiophenyl, pyrazolyl, benzimidazolyl, naphthyl, furanyl, pyrrolyl, and pyridyl aromatic rings.

Mimetics of acidic amino acids can be generated by substitution by, e.g., non-carboxylate amino acids while maintaining a negative charge; (phosphono)alanine; sulfated threonine. Carboxyl side groups (e.g., aspartyl or glutamyl) can also be selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as, e.g., 1-cyclohexyl-3(2-morpholinyl-(4-ethyl) carbodiimide or 1-ethyl-3(4-azonia- 4,4-dimetholpentyl) carbodiimide. Aspartyl or glutamyl can also be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions. Mimetics of basic amino acids can be generated by substitution with, e.g., (in addition to lysine and arginine) the amino acids ornithine, citrulline, or (guanidino)-acetic acid, or (guanidino)alkyl-acetic acid, where alkyl is defined above. Nitrile derivative (e.g., containing the CN-moiety in place of COOH) can be substituted for asparagine or glutamine. Asparaginyl and glutaminyl residues can be deaminated to the corresponding aspartyl or glutamyl residues. Arginine residue mimetics can be generated by reacting arginyl with, e.g., one or more conventional reagents, including, e.g., phenylglyoxal, 2,3-butanedione, 1,2-cyclo-hexanedione, or ninhydrin, in one aspect under alkaline conditions. Tyrosine residue mimetics can be generated by reacting tyrosyl with, e.g., aromatic diazonium compounds or tetranitromethane. N-acetylimidizol and tetranitromethane can be used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively. Cysteine residue mimetics can be generated by reacting cysteinyl residues with, e.g., alpha-haloacetates such as 2-chloroacetic acid or chloroacetamide and corresponding amines; to give carboxymethyl or carboxyamidomethyl derivatives. Cysteine residue mimetics can also be generated by reacting cysteinyl residues with, e.g., bromo-trifluoroacetone, alpha-bromo-beta-(5-imidozoyl) propionic acid; chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide; methyl 2-pyridyl disulfide; p-chloromercuribenzoate; 2-chloromercuri-4 nitrophenol; or, chloro-7-nitrobenzo-oxa-1,3-diazole. Lysine mimetics can be generated (and amino terminal residues can be altered) by reacting lysinyl with, e.g., succinic or other carboxylic acid anhydrides. Lysine and other alpha-amino-containing residue mimetics can also be generated by reaction with imidoesters, such as methyl picolinimidate, pyridoxal phosphate, pyridoxal, chloroborohydride, trinitro-

benzenesulfonic acid, O-methylisourea, 2,4, pentanedione, and transamidase-catalyzed reactions with glyoxylate. Mimetics of methionine can be generated by reaction with, e.g., methionine sulfoxide. Mimetics of proline include, e.g., pipecolic acid, thiazolidine carboxylic acid, 3- or 4- hydroxy proline, dehydroproline, 3- or 4-methylproline, or 3,3,-dimethylproline. Histidine residue mimetics can be generated by reacting histidyl with, e.g., diethylprocarbonate or para-bromophenacyl bromide. Other mimetics include, e.g., those generated by hydroxylation of proline and lysine; phosphorylation of the hydroxyl groups of seryl or threonyl residues; methylation of the alpha-amino groups of lysine, arginine and histidine; acetylation of the N-terminal amine; methylation of main chain amide residues or substitution with N-methyl amino acids; or amidation of C-terminal carboxyl groups.

A residue, e.g., an amino acid, of a polypeptide of the invention can also be replaced by an amino acid (or peptidomimetic residue) of the opposite chirality. Thus, any amino acid naturally occurring in the L-configuration (which can also be referred to as the R or S, depending upon the structure of the chemical entity) can be replaced with the amino acid of the same chemical structural type or a peptidomimetic, but of the opposite chirality, referred to as the D- amino acid, but also can be referred to as the R- or S- form.

The invention also provides methods for modifying the polypeptides of the invention by either natural processes, such as post-translational processing (e.g., phosphorylation, acylation, etc), or by chemical modification techniques, and the resulting modified polypeptides. Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In one aspect, the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide may have many types of modifications. Modifications of peptides or polypeptides of the invention include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization,

selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. See, e.g., Creighton, T.E., Proteins – Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983).

Solid-phase chemical peptide synthesis methods can also be used to synthesize the polypeptide or fragments of the invention. Such method have been known in the art since the early 1960's (Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154, 1963) (See also Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) and have recently been employed in commercially available laboratory peptide design and synthesis kits (Cambridge Research Biochemicals). Such commercially available laboratory kits have generally utilized the teachings of H. M. Geysen et al, Proc. Natl. Acad. Sci., USA, 81:3998 (1984) and provide for synthesizing peptides upon the tips of a multitude of "rods" or "pins" all of which are connected to a single plate. When such a system is utilized, a plate of rods or pins is inverted and inserted into a second plate of corresponding wells or reservoirs, which contain solutions for attaching or anchoring an appropriate amino acid to the pin's or rod's tips. By repeating such a process step, i.e., inverting and inserting the rod's and pin's tips into appropriate solutions, amino acids are built into desired peptides. In addition, a number of available FMOC peptide synthesis systems are available. For example, assembly of a polypeptide or fragment can be carried out on a solid support using an Applied Biosystems, Inc. model 431A™ automated peptide synthesizer. Such equipment provides ready access to the peptides of the invention, either by direct synthesis or by synthesis of a series of fragments that can be coupled using other known techniques.

The polypeptides of the invention include a polypeptide, enzyme, protein, e.g. structural or binding protein, in an active or inactive form. For example, the polypeptides of the invention include proproteins before "maturation" or processing of prepro sequences, e.g., by a proprotein-processing enzyme, such as a proprotein convertase to generate an "active" mature protein. The polypeptides of the invention include a polypeptide, enzyme, protein, e.g. structural or binding protein, inactive for other reasons, e.g., before "activation" by a post-translational processing event, e.g., an endo- or exo-peptidase or proteinase action, a phosphorylation event, an amidation, a glycosylation or a sulfation, a dimerization event, and the like. The polypeptides of the

invention include all active forms, including active subsequences, e.g., catalytic domains or active sites, of the enzyme.

The invention includes immobilized polypeptides, enzymes, proteins, e.g. structural or binding proteins, anti-polypeptides, anti-enzymes, anti-proteins, e.g. anti-structural or anti-binding proteins, antibodies and fragments thereof. The invention provides methods for inhibiting a polypeptide, enzyme, protein, e.g. structural or binding protein, activity, e.g., using dominant negative mutants or anti-polypeptide, anti-enzyme, anti-protein, e.g. anti-structural or anti-binding protein antibodies of the invention. The invention includes heterocomplexes, e.g., fusion proteins, heterodimers, etc., comprising the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention.

Polypeptides of the invention can have an enzyme, structural or binding activity under various conditions, e.g., extremes in pH and/or temperature, oxidizing agents, and the like. The invention provides methods leading to alternative a polypeptide, enzyme, protein, e.g. structural or binding protein, preparations with different catalytic efficiencies and stabilities, e.g., towards temperature, oxidizing agents and changing wash conditions. In one aspect, a polypeptide, enzyme, protein, e.g. structural or binding protein, variants can be produced using techniques of site-directed mutagenesis and/or random mutagenesis. In one aspect, directed evolution can be used to produce a great variety of a polypeptide, enzyme, protein, e.g. structural or binding protein, variants with alternative specificities and stability.

The proteins of the invention are also useful as research reagents to identify a polypeptide, enzyme, protein, e.g. structural or binding protein, modulators, e.g., activators or inhibitors of a polypeptide, enzyme, protein, e.g. structural or binding protein, activity. Briefly, test samples (compounds, broths, extracts, and the like) are added to a polypeptide, enzyme, protein, e.g. structural or binding protein, assays to determine their ability to inhibit substrate cleavage. Inhibitors identified in this way can be used in industry and research to reduce or prevent undesired proteolysis. As with a polypeptide, enzyme, protein, e.g. structural or binding protein, inhibitors can be combined to increase the spectrum of activity.

The enzymes of the invention are also useful as research reagents to digest proteins or in protein sequencing. For example, the polypeptide, enzyme, protein, e.g. structural or binding proteins may be used to break polypeptides into smaller fragments for sequencing using, e.g. an automated sequencer.

The invention also provides methods of discovering new a polypeptide, enzyme, protein, e.g. structural or binding protein, using the nucleic acids, polypeptides and antibodies of the invention. In one aspect, phagemid libraries are screened for expression-based discovery of a polypeptide, enzyme, protein, e.g. structural or binding protein, . In another aspect, lambda phage libraries are screened for expression-based discovery of a polypeptide, enzyme, protein, e.g. structural or binding protein, . Screening of the phage or phagemid libraries can allow the detection of toxic clones; improved access to substrate; reduced need for engineering a host, by-passing the potential for any bias resulting from mass excision of the library; and, faster growth at low clone densities. Screening of phage or phagemid libraries can be in liquid phase or in solid phase. In one aspect, the invention provides screening in liquid phase. This gives a greater flexibility in assay conditions; additional substrate flexibility; higher sensitivity for weak clones; and ease of automation over solid phase screening.

The invention provides screening methods using the proteins and nucleic acids of the invention and robotic automation to enable the execution of many thousands of biocatalytic reactions and screening assays in a short period of time, e.g., per day, as well as ensuring a high level of accuracy and reproducibility (see discussion of arrays, below). As a result, a library of derivative compounds can be produced in a matter of weeks. For further teachings on modification of molecules, including small molecules, see PCT/US94/09174.

In one aspect, polypeptides or fragments of the invention may be obtained through biochemical enrichment or purification procedures. The sequence of potentially homologous polypeptides or fragments may be determined by a polypeptide, enzyme, protein, e.g. structural or binding protein, assays, gel electrophoresis and/or microsequencing. The sequence of the prospective polypeptide or fragment of the invention can be compared to an exemplary polypeptide of the invention, or a fragment, e.g., comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 or more consecutive amino acids thereof using any of the programs described above.

Another aspect of the invention is an assay for identifying fragments or variants of the invention, which retain the enzymatic function of the polypeptides of the invention. For example the fragments or variants of said polypeptides, may be used to catalyze biochemical reactions (e.g., production of a nootkatone from a valencene), which indicate that the fragment or variant retains the enzymatic activity of a polypeptide of the invention.

An exemplary assay for determining if fragments of variants retain the enzymatic activity of the polypeptides of the invention includes the steps of: contacting the polypeptide fragment or variant with a substrate molecule under conditions which allow the polypeptide fragment or variant to function and detecting either a decrease in the level of substrate or an increase in the level of the specific reaction product of the reaction between the polypeptide and substrate.

The present invention exploits the unique catalytic properties of enzymes. Whereas the use of biocatalysts (i.e., purified or crude enzymes, non-living or living cells) in chemical transformations normally requires the identification of a particular biocatalyst that reacts with a specific starting compound, the present invention uses selected biocatalysts and reaction conditions that are specific for functional groups that are present in many starting compounds, such as small molecules. Each biocatalyst is specific for one functional group, or several related functional groups and can react with many starting compounds containing this functional group.

The biocatalytic reactions produce a population of derivatives from a single starting compound. These derivatives can be subjected to another round of biocatalytic reactions to produce a second population of derivative compounds. Thousands of variations of the original small molecule or compound can be produced with each iteration of biocatalytic derivatization.

Enzymes react at specific sites of a starting compound without affecting the rest of the molecule, a process which is very difficult to achieve using traditional chemical methods. This high degree of biocatalytic specificity provides the means to identify a single active compound within the library. The library is characterized by the series of biocatalytic reactions used to produce it, a so-called "biosynthetic history". Screening the library for biological activities and tracing the biosynthetic history identifies the specific reaction sequence producing the active compound. The reaction sequence is repeated and the structure of the synthesized compound determined. This mode of identification, unlike other synthesis and screening approaches, does not require immobilization technologies and compounds can be synthesized and tested free in solution using virtually any type of screening assay. It is important to note, that the high degree of specificity of enzyme reactions on functional groups allows for the "tracking" of specific enzymatic reactions that make up the biocatalytically produced library.

Many of the procedural steps are performed using robotic automation enabling the execution of many thousands of biocatalytic reactions and screening assays per day as

well as ensuring a high level of accuracy and reproducibility. As a result, a library of derivative compounds can be produced in a matter of weeks, which would take years to produce using current chemical methods.

In a particular aspect, the invention provides a method for modifying small molecules, comprising contacting a polypeptide encoded by a polynucleotide described herein or enzymatically active fragments thereof with a small molecule to produce a modified small molecule. A library of modified small molecules is tested to determine if a modified small molecule is present within the library, which exhibits a desired activity. A specific biocatalytic reaction which produces the modified small molecule of desired activity is identified by systematically eliminating each of the biocatalytic reactions used to produce a portion of the library and then testing the small molecules produced in the portion of the library for the presence or absence of the modified small molecule with the desired activity. The specific biocatalytic reactions which produce the modified small molecule of desired activity is optionally repeated. The biocatalytic reactions are conducted with a group of biocatalysts that react with distinct structural moieties found within the structure of a small molecule, each biocatalyst is specific for one structural moiety or a group of related structural moieties; and each biocatalyst reacts with many different small molecules which contain the distinct structural moiety.

*A polypeptide, enzyme, protein, e.g. structural or binding protein, signal sequences, prepro and catalytic domains*

The invention provides a polypeptide, enzyme, protein, e.g. structural or binding protein, signal sequences (e.g., signal peptides (SPs)), prepro domains and catalytic domains (CDs). The SPs, prepro domains and/or CDs of the invention can be isolated, synthetic or recombinant peptides or can be part of a fusion protein, e.g., as a heterologous domain in a chimeric protein. The invention provides nucleic acids encoding these catalytic domains (CDs), prepro domains and signal sequences (SPs, e.g., a peptide having a sequence comprising/ consisting of amino terminal residues of a polypeptide of the invention).

The invention provides isolated, synthetic or recombinant signal sequences (e.g., signal peptides) consisting of or comprising a sequence as set forth in residues 1 to 14, 1 to 15, 1 to 16, 1 to 17, 1 to 18, 1 to 19, 1 to 20, 1 to 21, 1 to 22, 1 to 23, 1 to 24, 1 to 25, 1 to 26, 1 to 27, 1 to 28, 1 to 28, 1 to 30, 1 to 31, 1 to 32, 1 to 33, 1 to 34, 1 to 35, 1 to 36, 1 to 37, 1 to 38, 1 to 40, 1 to 41, 1 to 42, 1 to 43, 1 to 44, 1 to 45, 1 to 46, or 1 to 47, or more, of a polypeptide of the invention, e.g., an exemplary polypeptide of the invention

including SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, SEQ ID NO:81, SEQ ID NO:83, SEQ ID NO:85, SEQ ID NO:87, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:97, SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, SEQ ID NO:105, and all polypeptides (amino acid sequences) disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:3 through SEQ ID NO:108,699. In one aspect, the invention provides signal sequences (leader peptides, leader sequences) comprising, or consisting of, the first 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 or more amino terminal residues of a polypeptide of the invention, or comprising, or consisting of, the signal sequences (leader peptides, leader sequences) set forth in Table 2.

In one aspect, novel a polypeptide, enzyme, protein, e.g. structural or binding protein, signal peptides are identified by a method referred to as SignalP. SignalP uses a combined neural network which recognizes both signal peptides and their cleavage sites. See, e.g., Nielsen (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering 10:1-6. For example, in alternative embodiments, the invention provides isolated, synthetic or recombinant signal sequences comprising/ consisting of the signal sequences set forth in Table 2, below, and polypeptides comprising these signal sequences. The polypeptide can be an enzyme, peptide or protein of the invention. For example, as an aid to reading Table 2, the invention provides an isolated, synthetic or recombinant signal sequence as set forth by residues 1 to 23 of SEQ ID NO:78,304 (encoded, e.g., by SEQ ID NO:78,303), i.e., having the sequence MAAALAATALASAAAASSFHITA. This can be determined by reading the second column for the first row "Signalp Cleavage Site"/ "Probability: 0.998 AA1: 23 AA2: 24", i.e., the cleavage of signal sequence takes place between amino acid 23 (AA23) and amino acid 24 (AA24) with a probability of 0.992 that this is the correct

cleavage site. Therefore, the signal sequence is predicted to be from the amino acid in position 1 of SEQ ID NO:78,304 up to and including the amino acid in position 23 of SEQ ID NO:78,304. This signal sequence, in one aspect, is encoded by a subsequence of SEQ ID NO:78,303.

Table 2

| SEQ ID NO: | Signalp Cleavage Site Probability: | Predicted Signal Sequence |
|---|---|---|
| 78303, 78304 | 0.998 AA1: 23 AA2: 24 Probability: | MAAALAATALASAAAASSFHITA |
| 20757, 20758 | 0.966 AA1: 22 AA2: 23 Probability: | MAAFRNSLLIATTSAFSATVIG |
| 56437, 56438 | 0.975 AA1: 22 AA2: 23 Probability: | MAAGFLAAVFFLAAGFFFAVFL |
| 65235, 65236 | 0.873 AA1: 21 AA2: 22 Probability: | MAAKMKTTAIPPLTAAAAINA |
| 50827, 50828 | 0.998 AA1: 22 AA2: 23 Probability: | MAAPQASSTGSLTLRMSPSASA |
| 35077, 35078 | 0.824 AA1: 18 AA2: 19 Probability: | MAARAISPTLPTSRAVSA |
| 72299, 72300 | 0.995 AA1: 17 AA2: 18 Probability: | MAASSIGICFGSSAVMA |
| 74291, 74292 | 0.996 AA1: 24 AA2: 25 Probability: | MAATGLFSAFAAGCLVMAAGLVFM |
| 27739, 27740 | 0.979 AA1: | MAAVSVAAVSVAAVRVAAVSIAAVSIA |

| | | |
|---|---|---|
| | 27 AA2: 28 | |
| | Probability: | |
| 51679, | 0.573 AA1: | |
| 51680 | 25 AA2: 26 | MADALYFGILLFLIILCCILAPVLA |
| | Probability: | |
| 27853, | 0.922 AA1: | |
| 27854 | 20 AA2: 21 | MADIIILGATALGAFGLCFA |
| | Probability: | |
| 19857, | 0.968 AA1: | |
| 19858 | 20 AA2: 21 | MADKAIIAMSGGVDSSVAAA |
| | Probability: | |
| 61283, | 0.536 AA1: | |
| 61284 | 25 AA2: 26 | MADMVLAPIAAMLIQMAISRSREYA |
| | Probability: | |
| 46291, | 0.958 AA1: | |
| 46292 | 20 AA2: 21 | MADRIRMNTLCLALVTLSQA |
| | Probability: | |
| 62683, | 1.000 AA1: | |
| 62684 | 27 AA2: 28 | MADVDAALASTAVLASAAALASAAALA |
| | Probability: | |
| 34963, | 0.905 AA1: | |
| 34964 | 23 AA2: 24 | MAEETAAAAKTAATAGTAGTAAA |
| | Probability: | |
| 39789, | 0.762 AA1: | |
| 39790 | 20 AA2: 21 | MAEFPAASLPPAAFSQVAYT |
| | Probability: | |
| 13901, | 0.752 AA1: | |
| 13902 | 24 AA2: 25 | MAEIIGRNVMVISTGMGGPSTAIA |
| | Probability: | |
| 46733, | 0.870 AA1: | |
| 46734 | 20 AA2: 21 | MAEKAIIAMSGGVDSSVAAA |
| | Probability: | |
| 48239, | 0.578 AA1: | |
| 48240 | 30 AA2: 31 | MAERARPCGFRHKRRRHTGLAAFYSGTVSA |

| | | |
|---|---|---|
| | Probability: | |
| 69847, | 0.613 AA1: | |
| 69848 | 24 AA2: 25 | MAERRMNRKWMAGFYICGTGQAIS |
| | Probability: | |
| 42141, | 0.610 AA1: | |
| 42142 | 24 AA2: 25 | MAETLPLTALSLGLPQFFLSPAFS |
| | Probability: | |
| 49655, | 0.744 AA1: | |
| 49656 | 25 AA2: 26 | MAETMLFLLAGISIIGLGIMLYLRA |
| | Probability: | |
| 75131, | 0.870 AA1: | |
| 75132 | 23 AA2: 24 | MAEYKITGFLLVAVMIFGLAACS |
| | Probability: | |
| 15879, | 0.697 AA1: | |
| 15880 | 30 AA2: 31 | MAEYRKIANSIGFSIIDSFAGCAAGAGASG |
| | Probability: | |
| 74935, | 0.999 AA1: | |
| 74936 | 26 AA2: 27 | MAFAPARPSAFIFAPILALTFSPALA |
| | Probability: | |
| 52679, | 0.902 AA1: | |
| 52680 | 23 AA2: 24 | MAFARVRTVCSSSMTMIFFFSLS |
| | Probability: | |
| 9773, | 0.985 AA1: | |
| 9774 | 20 AA2: 21 | MAFASRAFCLSCVARALASS |
| | Probability: | |
| 20785, | 0.970 AA1: | |
| 20786 | 21 AA2: 22 | MAFFYRLLSSPFPYASAVLLA |
| | Probability: | |
| 6639, | 0.869 AA1: | |
| 6640 | 21 AA2: 22 | MAFIIPFTIVLFALAVISLVS |
| | Probability: | |
| 77375, | 0.925 AA1: | |
| 77376 | 23 AA2: 24 | MAFLMGSGVSPNAAGLIVPGAAA |
| 51841, | Probability: | MAFRSRSSASRSIVSISLSIPIFCRA |

| | | |
|---|---|---|
| 51842 | 0.833 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 13567, | 0.820 AA1: | |
| 13568 | 27 AA2: 28 | MAFTLIPHMSSIWSLMAVIIVFTFVMS |
| | Probability: | |
| 4979, | 0.812 AA1: | |
| 4980 | 25 AA2: 26 | MAFWGRSNLKRIMVTVIGAGSTARA |
| | Probability: | |
| 46549, | 0.930 AA1: | |
| 46550 | 26 AA2: 27 | MAGGQKRGTVVAAGTAVAPWSHAQGA |
| | Probability: | |
| 33123, | 0.679 AA1: | |
| 33124 | 19 AA2: 20 | MAGPVVAGAVVLVAGSAIS |
| | Probability: | |
| 23489, | 0.990 AA1: | |
| 23490 | 16 AA2: 17 | MAGRVIVLLLLVAVLA |
| | Probability: | |
| 73871, | 0.982 AA1: | |
| 73872 | 18 AA2: 19 | MAGTFTASIAAVASGATA |
| | Probability: | |
| 38633, | 0.998 AA1: | |
| 38634 | 27 AA2: 28 | MAHISKIYRTLLSALLALLGFSCGSCG |
| | Probability: | |
| 78293, | 0.706 AA1: | |
| 78294 | 24 AA2: 25 | MAHKLLERAARPLYAAVATALALA |
| | Probability: | |
| 47825, | 0.588 AA1: | |
| 47826 | 24 AA2: 25 | MAHTSLRYGYSAATILLSSNNAGS |
| | Probability: | |
| 71557, | 0.821 AA1: | |
| 71558 | 23 AA2: 24 | MAIFPVLKSALTIMIPIFFAAFG |
| 66157, | Probability: | |
| 66158 | 0.998 AA1: | MAIFTAIIAAGVVFTVTVFAVLTVSA |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 1389, | 0.716 AA1: | |
| 1390 | 22 AA2: 23 | MAIISILSFSLLFMGCPNITET |
| | Probability: | |
| 24291, | 0.900 AA1: | |
| 24292 | 20 AA2: 21 | MAILVATFAIGYPVALLASA |
| | Probability: | |
| 10363, | 0.984 AA1: | |
| 10364 | 30 AA2: 31 | MAIQRKKLIMLAAIGVGGGMFLGFARILFS |
| | Probability: | |
| 3929, | 0.504 AA1: | |
| 3930 | 29 AA2: 30 | MAITRVYSVWVIGGAAVIAVFLSFFNKFS |
| | Probability: | |
| 4811, | 0.998 AA1: | |
| 4812 | 26 AA2: 27 | MAIXTAIIAAGVVFTVTVFAVLTVSA |
| | Probability: | |
| 52377, | 0.912 AA1: | |
| 52378 | 15 AA2: 16 | MAKAPAASAPTGAQA |
| | Probability: | |
| 2479, | 1.000 AA1: | |
| 2480 | 21 AA2: 22 | MAKFLLFLLIAVTTGISPAFA |
| | Probability: | |
| 28517, | 0.986 AA1: | |
| 28518 | 29 AA2: 30 | MAKIIKSARNRFVFAAVFAAIIAVSNFFQ |
| | Probability: | |
| 43427, | 0.534 AA1: | |
| 43428 | 16 AA2: 17 | MAKKKRKPSATPPAAA |
| | Probability: | |
| 3617, | 0.863 AA1: | |
| 3618 | 20 AA2: 21 | MAKNFLSLTMSMTFEPKAQA |
| | Probability: | |
| 72345, | 0.709 AA1: | |
| 72346 | 27 AA2: 28 | MAKNKNPMRSQFIIFSCVLFLVILVVG |

| | | |
|---|---|---|
| | Probability: | |
| 40395, | 0.960 AA1: | |
| 40396 | 23 AA2: 24 | MAKPIVMSIFILLAVILSGTRMG |
| | Probability: | |
| 31435, | 0.790 AA1: | |
| 31436 | 19 AA2: 20 | MAKTSAAAATADGAAPAGG |
| | Probability: | |
| 70245, | 0.992 AA1: | |
| 70246 | 22 AA2: 23 | MAKTTSIITVIITAIAAVSLNA |
| | Probability: | |
| 5865, | 0.993 AA1: | |
| 5866 | 22 AA2: 23 | MAKVKTNWFIVIMLWLAGIAAA |
| | Probability: | |
| 9711, | 0.660 AA1: | |
| 9712 | 23 AA2: 24 | MALACIRAGADALSLVNSFKAMA |
| | Probability: | |
| 31003, | 0.940 AA1: | |
| 31004 | 18 AA2: 19 | MALCLLGNFVLAFGVAMA |
| | Probability: | |
| 24789, | 0.999 AA1: | |
| 24790 | 26 AA2: 27 | MALFVSRLGFAGLSPCAPGTCGTALA |
| | Probability: | |
| 57745, | 0.589 AA1: | |
| 57746 | 28 AA2: 29 | MALIFFSGGSMALPKAVMLSSRNFNALA |
| | Probability: | |
| 73817, | 0.740 AA1: | |
| 73818 | 22 AA2: 23 | MALMILSTVGAYSIVTPILASA |
| | Probability: | |
| 63431, | 0.671 AA1: | |
| 63432 | 23 AA2: 24 | MALPPFSWVFFSCAAFKSSSFFS |
| | Probability: | |
| 44739, | 0.927 AA1: | |
| 44740 | 26 AA2: 27 | MALRSRSSASRSIVSISLSIPVFCRA |
| 29757, | Probability: | MALRSRSSASRSMVSISLSIPIFCRA |

| | | |
|---|---|---|
| 29758 | 0.832 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 55375, | 0.928 AA1: | |
| 55376 | 26 AA2: 27 | MALRSRSSDSRSMASMSLSIPVFCRA |
| | Probability: | |
| 63703, | 0.928 AA1: | |
| 63704 | 26 AA2: 27 | MALRSRSSDSRSMASMSLSIPVFCRA |
| | Probability: | |
| 44143, | 0.632 AA1: | |
| 44144 | 22 AA2: 23 | MALTGHLVFSTLSTGDSASAIA |
| | Probability: | |
| 22113, | 0.575 AA1: | |
| 22114 | 21 AA2: 22 | MALTTSLTAMSMPRLISMTLA |
| | Probability: | |
| 38735, | 0.982 AA1: | |
| 38736 | 22 AA2: 23 | MALWFLLLISLTWFTATFTCSA |
| | Probability: | |
| 38511, | 1.000 AA1: | |
| 38512 | 25 AA2: 26 | MAMARRSSALACAIFLSASARSVSS |
| | Probability: | |
| 29425, | 0.683 AA1: | |
| 29426 | 22 AA2: 23 | MAMFLIAFLAVIARSIALSIYQ |
| | Probability: | |
| 28889, | 1.000 AA1: | |
| 28890 | 25 AA2: 26 | MAMFPSFSRKLCLFVLLFSALTAGA |
| | Probability: | |
| 74279, | 0.792 AA1: | |
| 74280 | 27 AA2: 28 | MAMFSSPFWRTNGFFGFMAMMVPISMS |
| | Probability: | |
| 19977, | 0.701 AA1: | |
| 19978 | 18 AA2: 19 | MAMLAAPAVAPDGGAGAA |
| 47801, | Probability: | |
| 47802 | 0.989 AA1: | MAMRLYSRRQVVFFSILSCLLTAFIVIG |

| | | |
|---|---|---|
| | 28 AA2: 29 | |
| | Probability: | |
| 2541, | 0.593 AA1: | |
| 2542 | 19 AA2: 20 | MAMSGAAPVPTPPPMPATA |
| | Probability: | |
| 30017, | 0.958 AA1: | |
| 30018 | 20 AA2: 21 | MAMSLLPLATSLMAPATRLA |
| | Probability: | |
| 49881, | 0.700 AA1: | |
| 49882 | 24 AA2: 25 | MANAATIAGMAFANAFLGVCHSMA |
| | Probability: | |
| 20445, | 0.714 AA1: | |
| 20446 | 22 AA2: 23 | MANALTHGMGVLFGVASIPVLS |
| | Probability: | |
| 49539, | 0.999 AA1: | |
| 49540 | 24 AA2: 25 | MANKLRLFVPIALLMAFAAGFCIA |
| | Probability: | |
| 54757, | 0.789 AA1: | |
| 54758 | 21 AA2: 22 | MANLIRFFAAAPVISSGIVLG |
| | Probability: | |
| 9201, | 0.958 AA1: | |
| 9202 | 22 AA2: 23 | MANNYLKAGAALTLLILGITLA |
| | Probability: | |
| 25269, | 0.999 AA1: | |
| 25270 | 18 AA2: 19 | MANRKLLTVMLAVALVLA |
| | Probability: | |
| 48567, | 0.997 AA1: | |
| 48568 | 23 AA2: 24 | MANRKLLTVMLAVALVSAMTACK |
| | Probability: | |
| 55703, | 0.998 AA1: | |
| 55704 | 22 AA2: 23 | MANRYQKYVCSLSLTMSAIVLA |
| | Probability: | |
| 41111, | 0.635 AA1: | |
| 41112 | 22 AA2: 23 | MANYYFKTGVVLTLLILGITLA |

| | | |
|---|---|---|
| | Probability: | |
| 39399, | 0.539 AA1: | |
| 39400 | 24 AA2: 25 | MAPPTNITSSISEALSFASASAFL |
| | Probability: | |
| 59177, | 0.584 AA1: | |
| 59178 | 26 AA2: 27 | MAQFINAVFNGFSLTSIILLTSLGLA |
| | Probability: | |
| 7749, | 0.611 AA1: | |
| 7750 | 28 AA2: 29 | MAQITLIGMLVLGVVALVLIYFLASFFS |
| | Probability: | |
| 13613, | 0.613 AA1: | |
| 13614 | 28 AA2: 29 | MAQITLIGMLVLGVVALVLIYFLASFFS |
| | Probability: | |
| 43259, | 0.614 AA1: | |
| 43260 | 26 AA2: 27 | MAQKEYSVGYALFIACAAAIGGFLFG |
| | Probability: | |
| 32651, | 0.963 AA1: | |
| 32652 | 19 AA2: 20 | MAQKLPRAAILPAVSTSQA |
| | Probability: | |
| 21867, | 0.608 AA1: | |
| 21868 | 25 AA2: 26 | MARAETSKKSALPIIGIVLSFLIFS |
| | Probability: | |
| 1313, | 0.999 AA1: | |
| 1314 | 23 AA2: 24 | MARALASGSCSAWASISAATSSA |
| | Probability: | |
| 61543, | 0.996 AA1: | |
| 61544 | 24 AA2: 25 | MARALSGSLSGFFLTVIFAAFSTA |
| | Probability: | |
| 8077, | 0.555 AA1: | |
| 8078 | 19 AA2: 20 | MARASWVSLLSMTIFSKSA |
| | Probability: | |
| 27657, | 0.793 AA1: | |
| 27658 | 15 AA2: 16 | MARFTQFAVAAAAQA |
| 76359, | Probability: | MARGRNVQYVASSCLSLLLSARVSL |

| | | |
|---|---|---|
| 76360 | 0.914 AA1: | |
| | 25 AA2: 26 | |
| | Probability: | |
| 56497, | 0.794 AA1: | |
| 56498 | 23 AA2: 24 | MARIIKTAAIAFLSLLILHLVLR |
| | Probability: | |
| 46015, | 0.916 AA1: | |
| 46016 | 25 AA2: 26 | MARISKIYRILISALMALLGFSCRN |
| | Probability: | |
| 17039, | 0.986 AA1: | |
| 17040 | 22 AA2: 23 | MARITAASIAAVVVSLQLGAFS |
| | Probability: | |
| 32929, | 1.000 AA1: | |
| 32930 | 19 AA2: 20 | MARKILLIILIFTAFLAWA |
| | Probability: | |
| 74389, | 0.836 AA1: | |
| 74390 | 26 AA2: 27 | MARKKTPVRLYVAVVLLLITLTVAFV |
| | Probability: | |
| 41449, | 0.550 AA1: | |
| 41450 | 18 AA2: 19 | MARLAAAQAVLNNPPAKA |
| | Probability: | |
| 26665, | 0.985 AA1: | |
| 26666 | 17 AA2: 18 | MARMSAMTAAVRTAAIA |
| | Probability: | |
| 8085, | 0.982 AA1: | |
| 8086 | 28 AA2: 29 | MARNSAVSLSISILSIITLILTVSSCSG |
| | Probability: | |
| 58587, | 1.000 AA1: | |
| 58588 | 27 AA2: 28 | MARRKTMNFSKIAFILSTAICALFAQS |
| | Probability: | |
| 76903, | 0.606 AA1: | |
| 76904 | 19 AA2: 20 | MARRSTPLAFRLTITSAFT |
| 21201, | Probability: | |
| 21202 | 0.972 AA1: | MARVDEFTFLFLLFASLAFA |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 24641, | 0.859 AA1: | |
| 24642 | 18 AA2: 19 | MARWFAALTALALIVAYL |
| | Probability: | |
| 34661, | 0.541 AA1: | |
| 34662 | 16 AA2: 17 | MASEMAAISYTLLALA |
| | Probability: | |
| 10071, | 0.930 AA1: | |
| 10072 | 24 AA2: 25 | MASFGLELAGLSLNVSIVSAFVLA |
| | Probability: | |
| 42197, | 0.940 AA1: | |
| 42198 | 23 AA2: 24 | MASGLTFASAVVLTAVLTSDALT |
| | Probability: | |
| 52681, | 0.599 AA1: | |
| 52682 | 25 AA2: 26 | MASHKISSISTIIVILTVIMLSITS |
| | Probability: | |
| 77579, | 0.778 AA1: | |
| 77580 | 19 AA2: 20 | MASILSNASVVLIISTALS |
| | Probability: | |
| 33837, | 0.996 AA1: | |
| 33838 | 29 AA2: 30 | MASLASFASRAALASRASLTSLASFNSLA |
| | Probability: | |
| 9639, | 0.997 AA1: | |
| 9640 | 23 AA2: 24 | MASLASLASXASRASLASRASLA |
| | Probability: | |
| 73561, | 0.913 AA1: | |
| 73562 | 21 AA2: 22 | MASLTFTVLLSLFMISAGEQK |
| | Probability: | |
| 76877, | 0.759 AA1: | |
| 76878 | 23 AA2: 24 | MASNAINAVWGTLFLLSVLNVMG |
| | Probability: | |
| 21629, | 0.986 AA1: | |
| 21630 | 24 AA2: 25 | MASRIILFSRYFGIFSLFAAGVFA |

| | | |
|---|---|---|
| | Probability: | |
| 42031, | 0.913 AA1: | |
| 42032 | 17 AA2: 18 | MASRSAVMSAISFSSLA |
| | Probability: | |
| 61285, | 0.994 AA1: | |
| 61286 | 26 AA2: 27 | MASSIRSKILIPCVGIAVVCVVASTA |
| | Probability: | |
| 74479, | 0.862 AA1: | |
| 74480 | 21 AA2: 22 | MASSVSAFIIISLLVFPSRIA |
| | Probability: | |
| 41081, | 0.976 AA1: | |
| 41082 | 25 AA2: 26 | MASVNEMLVMTVIISGIVLAVASAA |
| | Probability: | |
| 43823, | 0.569 AA1: | |
| 43824 | 27 AA2: 28 | MASVSIRLKTIFILIATVVVITVSTLG |
| | Probability: | |
| 48593, | 0.550 AA1: | |
| 48594 | 26 AA2: 27 | MATAVSGVSLVDTVNFLFITAFNALA |
| | Probability: | |
| 40739, | 0.818 AA1: | |
| 40740 | 20 AA2: 21 | MATMIFLFLPLVILILYSFN |
| | Probability: | |
| 11437, | 0.833 AA1: | |
| 11438 | 25 AA2: 26 | MATRAASRGASLIKRSASASRAARA |
| | Probability: | |
| 8053, | 0.965 AA1: | |
| 8054 | 20 AA2: 21 | MAVELAVAMGVTIESQAAFA |
| | Probability: | |
| 69143, | 0.802 AA1: | |
| 69144 | 17 AA2: 18 | MAVGIGLFIALIGFANA |
| | Probability: | |
| 47659, | 0.868 AA1: | |
| 47660 | 23 AA2: 24 | MAVKLLSPVLHSAVTIMTPILLA |
| 30285, | Probability: | MAVRFWAVISLTLFTAVFTCSAPLAIS |

| | | |
|---|---|---|
| 30286 | 0.992 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 8567, | 0.995 AA1: | |
| 8568 | 23 AA2: 24 | MAVRLPALAVARQLIALSSGAIA |
| | Probability: | |
| 60617, | 0.503 AA1: | |
| 60618 | 24 AA2: 25 | MAVSPSNLSLCSGRSRCISGSALS |
| | Probability: | |
| 23325, | 0.675 AA1: | |
| 23326 | 26 AA2: 27 | MAWQLSILGLTFTLLLSSILIRIINA |
| | Probability: | |
| 22351, | 0.979 AA1: | |
| 22352 | 26 AA2: 27 | MAWRHNMFRKLFFLCVSFLVVESLFA |
| | Probability: | |
| 4623, | 0.626 AA1: | |
| 4624 | 23 AA2: 24 | MAYVISDSCTGCTACARSCPVFA |
| | Probability: | |
| 26145, | 0.934 AA1: | |
| 26146 | 19 AA2: 20 | MCAVAVLFIAALLAFSVMV |
| | Probability: | |
| 61281, | 0.726 AA1: | |
| 61282 | 25 AA2: 26 | MCAVMNFIKTTMLLATLTAIVVAFG |
| | Probability: | |
| 49155, | 0.997 AA1: | |
| 49156 | 28 AA2: 29 | MCIFAKRTTAILWVAVVLCLTFASVPRA |
| | Probability: | |
| 65339, | 1.000 AA1: | |
| 65340 | 25 AA2: 26 | MCQNMNKHPYSCLLLILLCSAPVFA |
| | Probability: | |
| 45099, | 0.685 AA1: | |
| 45100 | 20 AA2: 21 | MCQRVMIAMALSCNPEILIA |
| 45569, | Probability: | |
| 45570 | 0.646 AA1: | MCRCCDVLLSFPSLLLAFLFVTALG |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 61583, | 0.988 AA1: | |
| 61584 | 24 AA2: 25 | MCTSMIRPSAPAAAAAIESGYTRA |
| | Probability: | |
| 5019, | 0.850 AA1: | |
| 5020 | 20 AA2: 21 | MCVPTGRNSAFFASAHTASA |
| | Probability: | |
| 73751, | 0.526 AA1: | |
| 73752 | 18 AA2: 19 | MCXYSSLFMLXMSAILMV |
| | Probability: | |
| 435, | 0.507 AA1: | |
| 436 | 16 AA2: 17 | MDALYLLCGLASPSSA |
| | Probability: | |
| 76883, | 0.984 AA1: | |
| 76884 | 28 AA2: 29 | MDCLFQFMRRLAVVAVTAVTVICLPSYG |
| | Probability: | |
| 20695, | 0.997 AA1: | |
| 20696 | 28 AA2: 29 | MDDSNSVNIILLSILFILMCFSAFFSLA |
| | Probability: | |
| 24359, | 0.858 AA1: | |
| 24360 | 19 AA2: 20 | MDEGLALAALDVAGAPAQA |
| | Probability: | |
| 71215, | 1.000 AA1: | |
| 71216 | 27 AA2: 28 | MDGKRLFARVTLLASVVLLGLSTAASA |
| | Probability: | |
| 4591, | 0.937 AA1: | |
| 4592 | 26 AA2: 27 | MDKEKKKLLLVAVSVGVFLLVTITVA |
| | Probability: | |
| 37191, | 0.785 AA1: | |
| 37192 | 19 AA2: 20 | MDKKALIAMSGGVDSSVAA |
| | Probability: | |
| 4727, | 0.919 AA1: | |
| 4728 | 27 AA2: 28 | MDKNPMERTLLVCCGTGCIANGAKAVA |

| | | |
|---|---|---|
| | Probability: | |
| 11835, | 0.741 AA1: | |
| 11836 | 20 AA2: 21 | MDKRIPPGSGLGGGSSDAAA |
| | Probability: | |
| 30491, | 0.739 AA1: | |
| 30492 | 27 AA2: 28 | MDLFPLFNSLRISAIATSAAFVLGILA |
| | Probability: | |
| 24109, | 0.888 AA1: | |
| 24110 | 25 AA2: 26 | MDLKEASVILAMVFMFIGGGSGSTA |
| | Probability: | |
| 21841, | 0.569 AA1: | |
| 21842 | 20 AA2: 21 | MDLSNAISITAALIGFYFFG |
| | Probability: | |
| 35593, | 0.653 AA1: | |
| 35594 | 29 AA2: 30 | MDPKVLGMIGAGVVMGIAALGSGVGIGIA |
| | Probability: | |
| 31695, | 0.764 AA1: | |
| 31696 | 23 AA2: 24 | MDPRILGMIGAGIVMGIAALGSA |
| | Probability: | |
| 47211, | 0.609 AA1: | |
| 47212 | 19 AA2: 20 | MDPRTKLVLTALFTVSVFV |
| | Probability: | |
| 23383, | 0.989 AA1: | |
| 23384 | 24 AA2: 25 | MDRCCKKNPQPCFCLALSVSALFA |
| | Probability: | |
| 20763, | 0.831 AA1: | |
| 20764 | 28 AA2: 29 | MDRFGRVFCPGLGVSSSSTILGLSSAFS |
| | Probability: | |
| 9891, | 1.000 AA1: | |
| 9892 | 23 AA2: 24 | MDRIIRAVAAAALVAAFAGSAVA |
| | Probability: | |
| 11363, | 0.943 AA1: | |
| 11364 | 21 AA2: 22 | MDRLFSVILTLTVAVFFTAGA |
| 9395, | Probability: | MDRYGVAPKKIVYAIILCIIIFSGCASVKS |

| | | |
|---|---|---|
| 9396 | 0.758 AA1: | |
| | 30 AA2: 31 | |
| | Probability: | |
| 56459, | 0.693 AA1: | |
| 56460 | 26 AA2: 27 | MDSKCAVVSVVKSPALWLSLAVGAFA |
| | Probability: | |
| 55063, | 0.928 AA1: | |
| 55064 | 19 AA2: 20 | MDVIFACGGAVGNSVMAAA |
| | Probability: | |
| 57783, | 1.000 AA1: | |
| 57784 | 25 AA2: 26 | MDYRSFFAVMFAVLSTLITPTAAQA |
| | Probability: | |
| 56733, | 0.503 AA1: | |
| 56734 | 20 AA2: 21 | MEAAVKRMSGSIAASLPFSA |
| | Probability: | |
| 13249, | 0.519 AA1: | |
| 13250 | 20 AA2: 21 | MEASAATALVTGAFRGIAKA |
| | Probability: | |
| 4239, | 0.997 AA1: | |
| 4240 | 24 AA2: 25 | MEASMSRKLLFVCVALLAVESAFA |
| | Probability: | |
| 53239, | 0.682 AA1: | |
| 53240 | 24 AA2: 25 | MEAYMCRKLLFLCVSFFITESLFA |
| | Probability: | |
| 58263, | 0.998 AA1: | |
| 58264 | 24 AA2: 25 | MEAYMCRKLLFLCVSFFVAGSLFA |
| | Probability: | |
| 14921, | 0.963 AA1: | |
| 14922 | 26 AA2: 27 | MEEKLKRNILLFAVLAALVMTGISCA |
| | Probability: | |
| 40737, | 0.746 AA1: | |
| 40738 | 26 AA2: 27 | MEEKLKRSFVLFTALTVLGMIGISCS |
| 62783, | Probability: | |
| 62784 | 0.930 AA1: | MEENLINQRKFISALVLAALFAALIS |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 44713, | 0.840 AA1: | |
| 44714 | 29 AA2: 30 | MEFKVNKKKFVTFIIILLIILFALFIVFS |
| | Probability: | |
| | 0.995 AA1: | |
| 37, 38 | 24 AA2: 25 | MEFQMKKINMLIFLVFIFSLDTFA |
| | Probability: | |
| 36089, | 0.993 AA1: | |
| 36090 | 26 AA2: 27 | MEFSSFDTMWVLIAAALVFFMQAGFA |
| | Probability: | |
| 3449, | 0.672 AA1: | |
| 3450 | 28 AA2: 29 | MEGSKMKPTNRTHKQSAKAASICSAVLA |
| | Probability: | |
| 3677, | 0.731 AA1: | |
| 3678 | 19 AA2: 20 | MEIKSLVPIALLIASNVFM |
| | Probability: | |
| 71993, | 0.839 AA1: | |
| 71994 | 25 AA2: 26 | MEILKLLSIGLALGILVVIPGLSTG |
| | Probability: | |
| 49957, | 0.992 AA1: | |
| 49958 | 28 AA2: 29 | MEILKLLSIGLALGILVVIPGLSTGTMA |
| | Probability: | |
| 34989, | 0.963 AA1: | |
| 34990 | 18 AA2: 19 | MEITLYFSGAGCSLPAQA |
| | Probability: | |
| 25663, | 0.997 AA1: | |
| 25664 | 28 AA2: 29 | MEKEIMKAQFHKIALAAALTLATALALS |
| | Probability: | |
| 43317, | 0.988 AA1: | |
| 43318 | 26 AA2: 27 | MEKIKLSLGILAVTGLVSCASGPRIS |
| | Probability: | |
| 59423, | 0.998 AA1: | |
| 59424 | 23 AA2: 24 | MEKKKLLLVAISVGVFLALTIGA |

| | | |
|---|---|---|
| | Probability: | |
| 42787, | 0.891 AA1: | |
| 42788 | 25 AA2: 26 | MEKKMRKQSVTNRIISVLICCFLSA |
| | Probability: | |
| 61357, | 0.523 AA1: | |
| 61358 | 25 AA2: 26 | MEKNAALAFRSAAVFIILFQIRLIA |
| | Probability: | |
| 48215, | 0.999 AA1: | |
| 48216 | 25 AA2: 26 | MEKPMNRQFWCFIPAVALSLSTAFA |
| | Probability: | |
| 72363, | 0.731 AA1: | |
| 72364 | 25 AA2: 26 | MEKRFAVAFRSAAVFLILFQIRLIA |
| | Probability: | |
| 10347, | 0.568 AA1: | |
| 10348 | 24 AA2: 25 | MEKRKKKKINIVSLIINLFSTIIA |
| | Probability: | |
| 14505, | 0.912 AA1: | |
| 14506 | 20 AA2: 21 | MEKSKRLYIIGAGFAGRTLA |
| | Probability: | |
| 24831, | 0.886 AA1: | |
| 24832 | 20 AA2: 21 | MEKSKRLYIIGAGFAGRTLA |
| | Probability: | |
| 70269, | 0.887 AA1: | |
| 70270 | 20 AA2: 21 | MEKSKRLYIIGAGFAGRTLA |
| | Probability: | |
| 423, | 0.987 AA1: | |
| 424 | 24 AA2: 25 | MELGVMILIAALILLAIIIFASVA |
| | Probability: | |
| 28541, | 0.518 AA1: | |
| 28542 | 17 AA2: 18 | MELLAVLVPGIYFCAFA |
| | Probability: | |
| 62309, | 0.914 AA1: | |
| 62310 | 27 AA2: 28 | MELTSAARAASCTIASVVNPSSMPSVA |
| 78483, | Probability: | MEMLPNLMAVIAASFVGTVTA |

| | | |
|---|---|---|
| 78484 | 0.994 AA1: 21 AA2: 22 Probability: | |
| 53511, 53512 | 0.917 AA1: 28 AA2: 29 Probability: | MEMTNEKKVILIASVFLLIMVIITFSCA |
| 2857, 2858 | 0.785 AA1: 29 AA2: 30 Probability: | MENLSAALTVFALIIAGGFFSVFDSALCA |
| 34597, 34598 | 0.588 AA1: 21 AA2: 22 Probability: | MENMNLLKEASLTVGSALLFA |
| 47757, 47758 | 0.664 AA1: 17 AA2: 18 Probability: | MENMRSMAALLSTNTFS |
| 67111, 67112 | 0.771 AA1: 18 AA2: 19 Probability: | MENPFFSPSSSALAWAFA |
| 46535, 46536 | 0.838 AA1: 26 AA2: 27 Probability: | MEPFPAFFIVLALILFSGLFSLFDYA |
| 42449, 42450 | 0.532 AA1: 25 AA2: 26 Probability: | MEPVYEMITITIVISAVVLALVVAA |
| 8411, 8412 | 0.995 AA1: 18 AA2: 19 Probability: | MERFSFFFAILGSFLVFA |
| 34771, 34772 | 0.612 AA1: 27 AA2: 28 Probability: | MERGIHSGLYREGAWAIAAFLAALAGA |
| 9105, 9106 | 0.535 AA1: 30 AA2: 31 Probability: | MESKFSGGVGAIFLFCIWAPILSIITLGLA |
| 51559, 51560 | 1.000 AA1: | MESNMFRKLLFFCVLFLAVVSLFA |

| | | |
|---|---|---|
| | 24 AA2: 25 | |
| | Probability: | |
| 6219, | 0.778 AA1: | |
| 6220 | 25 AA2: 26 | METPMPIYAKAIPLILAGVLLNAFA |
| | Probability: | |
| 53571, | 0.524 AA1: | |
| 53572 | 18 AA2: 19 | METSAIPVLVLFGPTASG |
| | Probability: | |
| 60723, | 0.527 AA1: | |
| 60724 | 18 AA2: 19 | METSAIPVLVLFGPTASG |
| | Probability: | |
| 11257, | 0.744 AA1: | |
| 11258 | 28 AA2: 29 | METSFSKIDMKLVIPLFTGVLCSLLACA |
| | Probability: | |
| 64475, | 0.737 AA1: | |
| 64476 | 28 AA2: 29 | METSFSKINLKFIMPMFTGALCSMLVCA |
| | Probability: | |
| 71267, | 0.928 AA1: | |
| 71268 | 24 AA2: 25 | MEVAMKNLYFLLVLVFALSGCSNT |
| | Probability: | |
| 77857, | 0.650 AA1: | |
| 77858 | 26 AA2: 27 | MEVFMIRKTLAVIFTFVVVLTGCSRA |
| | Probability: | |
| 50119, | 0.960 AA1: | |
| 50120 | 24 AA2: 25 | MEVFMKHLFAFFVVLVICSCAALT |
| | Probability: | |
| 53397, | 0.713 AA1: | |
| 53398 | 25 AA2: 26 | MEVFMKHLFAFFVVVVICGCAALTA |
| | Probability: | |
| 27131, | 0.998 AA1: | |
| 27132 | 23 AA2: 24 | MEVFMKNICLFCLLCIFISNSYS |
| | Probability: | |
| 53163, | 0.625 AA1: | |
| 53164 | 17 AA2: 18 | MEVLMALLPIVVIAAFA |

| | | |
|---|---|---|
| | Probability: | |
| 4179, | 0.514 AA1: | |
| 4180 | 19 AA2: 20 | MEVSLGGGLAILSPPGILA |
| | Probability: | |
| 70717, | 0.502 AA1: | |
| 70718 | 23 AA2: 24 | MEVVHALFLSSPPLSTYSGSVLS |
| | Probability: | |
| 29121, | 0.825 AA1: | |
| 29122 | 24 AA2: 25 | MFAFPNAAKAAASNLYAMAFVVYA |
| | Probability: | |
| 65107, | 0.927 AA1: | |
| 65108 | 28 AA2: 29 | MFAFPNKTGKHFSVLALLIFSLSFPYSV |
| | Probability: | |
| 53083, | 0.821 AA1: | |
| 53084 | 18 AA2: 19 | MFALAFALCALFTLVGCA |
| | Probability: | |
| 42673, | 0.985 AA1: | |
| 42674 | 18 AA2: 19 | MFCLMALFCLVLPLSAQN |
| | Probability: | |
| 41193, | 0.537 AA1: | |
| 41194 | 18 AA2: 19 | MFCMGIGMTAMPLICFLA |
| | Probability: | |
| 38855, | 0.901 AA1: | |
| 38856 | 28 AA2: 29 | MFFAILLVLFLLVCVFICLLILIQSDKG |
| | Probability: | |
| 393, | 0.952 AA1: | |
| 394 | 28 AA2: 29 | MFFAILLVLFLLVCVFLCLLILIQSDKG |
| | Probability: | |
| 70513, | 0.529 AA1: | |
| 70514 | 21 AA2: 22 | MFFDQYYLLLVVPTLILSIWA |
| | Probability: | |
| 48893, | 0.935 AA1: | |
| 48894 | 20 AA2: 21 | MFFFLFSLFFLLFSFVLCYA |
| 68505, | Probability: | MFFLCVLCVLSASVRVSFVYA |

| | | |
|---|---|---|
| 68506 | 1.000 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| 5693, | 1.000 AA1: | |
| 5694 | 18 AA2: 19 | MFFRISLFLLFLCVSAFA |
| | Probability: | |
| 65601, | 1.000 AA1: | |
| 65602 | 23 AA2: 24 | MFFRRVVLCLLGNFVLAFGVAMA |
| | Probability: | |
| 70257, | 0.997 AA1: | |
| 70258 | 24 AA2: 25 | MFFSSVMLLLFCFFCMISLRNIMA |
| | Probability: | |
| 6969, | 0.950 AA1: | |
| 6970 | 25 AA2: 26 | MFFTGLYFTGSVTAAVXSAILTAIA |
| | Probability: | |
| 19533, | 0.941 AA1: | |
| 19534 | 24 AA2: 25 | MFFTIAIIFFYVMIAVSAYFFLTA |
| | Probability: | |
| 72935, | 0.998 AA1: | |
| 72936 | 20 AA2: 21 | MFFVVVALLLLFLQTGACSA |
| | Probability: | |
| 13443, | 0.633 AA1: | |
| 13444 | 21 AA2: 22 | MFFWGLMSLASFGFPGTNGFV |
| | Probability: | |
| 56533, | 0.857 AA1: | |
| 56534 | 24 AA2: 25 | MFFWRNRMKIFLTGGTGFIGHFVA |
| | Probability: | |
| 23701, | 0.897 AA1: | |
| 23702 | 26 AA2: 27 | MFGFDKPGNSLGLTLTGSTLAGPVWA |
| | Probability: | |
| 4139, | 0.998 AA1: | |
| 4140 | 21 AA2: 22 | MFGGAAAATAAAASKALPTSA |
| 17189, | Probability: | |
| 17190 | 0.667 AA1: | MFGILGSFLWYGASYSSTMYAMA |

| | | |
|---|---|---|
| | 23 AA2: 24 | |
| | Probability: | |
| 28671, | 0.601 AA1: | |
| 28672 | 25 AA2: 26 | MFGLIPASLIGIASPLCMYGTIPIA |
| | Probability: | |
| 65625, | 0.964 AA1: | |
| 65626 | 25 AA2: 26 | MFGWVPALLWLLFGSVFFGAVHDFA |
| | Probability: | |
| 33661, | 0.856 AA1: | |
| 33662 | 25 AA2: 26 | MFGWVPVLLWILVGGVFFGAVQDFA |
| | Probability: | |
| 66083, | 0.899 AA1: | |
| 66084 | 19 AA2: 20 | MFHAERWAGAFLSVSGANA |
| | Probability: | |
| 8063, | 0.584 AA1: | |
| 8064 | 30 AA2: 31 | MFHAERWAGAFLSVSGETANEAFMCLKALA |
| | Probability: | |
| 57875, | 0.846 AA1: | |
| 57876 | 19 AA2: 20 | MFHAERWAGAFLTVSGANA |
| | Probability: | |
| 5937, | 0.998 AA1: | |
| 5938 | 30 AA2: 31 | MFHRVLALPIFAAVMFIVYFVSVTTVGSLA |
| | Probability: | |
| 30909, | 0.714 AA1: | |
| 30910 | 27 AA2: 28 | MFIDIIVFALLGVFLTLGVMQGFIVSA |
| | Probability: | |
| 52927, | 0.976 AA1: | |
| 52928 | 19 AA2: 20 | MFIFCILLNFLVLTQNAQG |
| | Probability: | |
| 13463, | 0.986 AA1: | |
| 13464 | 26 AA2: 27 | MFILCFLSRLVKKSIAFGSIMAASIA |
| | Probability: | |
| 75619, | 0.979 AA1: | |
| 75620 | 30 AA2: 31 | MFILIMTTFFRNSALLAISVLAVIISISCS |

| | | |
|---|---|---|
| | Probability: | |
| 62993, | 0.692 AA1: | |
| 62994 | 26 AA2: 27 | MFIMRFFTRRGWTLLLTAAVIIGAIS |
| | Probability: | |
| 45529, | 1.000 AA1: | |
| 45530 | 23 AA2: 24 | MFIRRMLLSLLGNFVLAFGVAVA |
| | Probability: | |
| 64507, | 0.752 AA1: | |
| 64508 | 25 AA2: 26 | MFIRSQFTPFKAIFLFFLTSLWGYA |
| | Probability: | |
| 47133, | 0.999 AA1: | |
| 47134 | 27 AA2: 28 | MFIWYSNTMKKILFTVLIAALCLTACA |
| | Probability: | |
| 68153, | 0.999 AA1: | |
| 68154 | 27 AA2: 28 | MFKELFMKKVFFCLALILSVLLFFACA |
| | Probability: | |
| 25461, | 0.978 AA1: | |
| 25462 | 28 AA2: 29 | MFKFKLNKNIFACCLIYFVIVSCSIVRN |
| | Probability: | |
| 15441, | 0.999 AA1: | |
| 15442 | 20 AA2: 21 | MFKKFTWLFVYXAISSLTWA |
| | Probability: | |
| 52389, | 0.554 AA1: | |
| 52390 | 19 AA2: 20 | MFKKIPHTYAIIFGLIVCA |
| | Probability: | |
| 21605, | 0.999 AA1: | |
| 21606 | 18 AA2: 19 | MFKKSFLLILCVFSAVLA |
| | Probability: | |
| 26465, | 0.996 AA1: | |
| 26466 | 20 AA2: 21 | MFKLYHVLYASALLFASVYA |
| | Probability: | |
| 34931, | 1.000 AA1: | |
| 34932 | 25 AA2: 26 | MFKSIFSVMMPLLFAASGGLFPALA |
| 22513, | Probability: | MFLAALSFGMLVGRFIYVASDGLVSSTLLA |

| | | |
|---|---|---|
| 22514 | 0.845 AA1: | |
| | 30 AA2: 31 | |
| | Probability: | |
| 46037, | 0.629 AA1: | |
| 46038 | 22 AA2: 23 | MFLAFSLITLGGAISLLMQRNA |
| | Probability: | |
| 52863, | 1.000 AA1: | |
| 52864 | 27 AA2: 28 | MFLAYNALRRVRALALVLLFASATASQ |
| | Probability: | |
| 31869, | 0.986 AA1: | |
| 31870 | 19 AA2: 20 | MFLLGVALFGTSSVACAVA |
| | Probability: | |
| 69429, | 0.955 AA1: | |
| 69430 | 19 AA2: 20 | MFLRSLSVPVSAISGIGVS |
| | Probability: | |
| 40623, | 0.990 AA1: | |
| 40624 | 25 AA2: 26 | MFLRTKEFIKKPFLLLPVLLFFACA |
| | Probability: | |
| 14405, | 0.827 AA1: | |
| 14406 | 24 AA2: 25 | MFLVSVFCFLAFSVIRGDTASFLA |
| | Probability: | |
| 15021, | 0.796 AA1: | |
| 15022 | 28 AA2: 29 | MFLVVSCPCALVISIPLSFFGGIGGASR |
| | Probability: | |
| 38799, | 0.903 AA1: | |
| 38800 | 25 AA2: 26 | MFMRFHFVPLRAVVLFFLTSLCGYA |
| | Probability: | |
| 71149, | 0.542 AA1: | |
| 71150 | 18 AA2: 19 | MFMVMSPSFVSIHAPARG |
| | Probability: | |
| 39103, | 0.851 AA1: | |
| 39104 | 27 AA2: 28 | MFNLSKKWVFITGAARGIGKQAALAFA |
| 42911, | Probability: | |
| 42912 | 0.945 AA1: | MFNYLAALAVTGNSAAGFCSFA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 36679, | 0.995 AA1: | |
| 36680 | 22 AA2: 23 | MFPAFLLCSSGVSSARIGSALA |
| | Probability: | |
| 76861, | 0.828 AA1: | |
| 76862 | 21 AA2: 22 | MFPDSRCPILLALMLTEFSHA |
| | Probability: | |
| 11185, | 0.625 AA1: | |
| 11186 | 24 AA2: 25 | MFPETSGAVVSIAAIKGVTPPAAG |
| | Probability: | |
| 19347, | 0.585 AA1: | |
| 19348 | 16 AA2: 17 | MFPISLTLSAVYVAVS |
| | Probability: | |
| 50685, | 0.881 AA1: | |
| 50686 | 25 AA2: 26 | MFPLKFKYLIIAFSILIVLFLLITA |
| | Probability: | |
| 65093, | 0.999 AA1: | |
| 65094 | 28 AA2: 29 | MFPVTHKTGKFFPVLALLIFSLCFPDIA |
| | Probability: | |
| 21103, | 0.737 AA1: | |
| 21104 | 28 AA2: 29 | MFRAVVYMPNLLTAVSVALLFRSXXGFG |
| | Probability: | |
| 26015, | 0.999 AA1: | |
| 26016 | 17 AA2: 18 | MFRFLIFILISAIFAFA |
| | Probability: | |
| 6671, | 0.568 AA1: | |
| 6672 | 19 AA2: 20 | MFRFWAILSGGFSSMKEFG |
| | Probability: | |
| 35461, | 0.940 AA1: | |
| 35462 | 19 AA2: 20 | MFRIAPLMRFMLNSVSVSA |
| | Probability: | |
| 60159, | 0.999 AA1: | |
| 60160 | 20 AA2: 21 | MFRKLLFLCVLLLIVETSFA |

| | | |
|---|---|---|
| | Probability: | |
| 21937, | 0.956 AA1: | |
| 21938 | 21 AA2: 22 | MFRLFLIDSSALAFRMFYAYA |
| | Probability: | |
| 69261, | 0.957 AA1: | |
| 69262 | 21 AA2: 22 | MFRLFLIDSSALAFRMFYAYA |
| | Probability: | |
| 67183, | 1.000 AA1: | |
| 67184 | 24 AA2: 25 | MFRLLRVHFFFFLSLLVFSCASQA |
| | Probability: | |
| 32267, | 0.986 AA1: | |
| 32268 | 18 AA2: 19 | MFRLTPATSPVVTVTVHA |
| | Probability: | |
| 17723, | 0.632 AA1: | |
| 17724 | 28 AA2: 29 | MFRSESMNKSLNMGSLIGALGAMFMIVG |
| | Probability: | |
| 74685, | 0.908 AA1: | |
| 74686 | 25 AA2: 26 | MFSIILLLVPIILSVISIPLMSLLG |
| | Probability: | |
| 33581, | 0.617 AA1: | |
| 33582 | 16 AA2: 17 | MFSKGGFVSVPPCAAA |
| | Probability: | |
| 17619, | 0.742 AA1: | |
| 17620 | 16 AA2: 17 | MFSKGGLXSVPPCAAA |
| | Probability: | |
| 27673, | 0.618 AA1: | |
| 27674 | 29 AA2: 30 | MFSLFLKKSIKSKKNENSNFVSIIISAHN |
| | Probability: | |
| 11299, | 1.000 AA1: | |
| 11300 | 23 AA2: 24 | MFSLKRFFILIITAVIIVSTAAA |
| | Probability: | |
| 39697, | 0.916 AA1: | |
| 39698 | 29 AA2: 30 | MFSLRYRRRPSRAGGAAATFTVRPQSGLS |
| 2253, | Probability: | MFSMKRFSILILTGVIAVSAVSA |

| | | |
|---|---|---|
| 2254 | 1.000 AA1: | |
| | 23 AA2: 24 | |
| | Probability: | |
| 31287, | 0.963 AA1: | |
| 31288 | 22 AA2: 23 | MFSNGIIILAIVSSFLVWLLHA |
| | Probability: | |
| 1693, | 1.000 AA1: | |
| 1694 | 20 AA2: 21 | MFSTILFILFAAVCLASAFA |
| | Probability: | |
| 55005, | 0.526 AA1: | |
| 55006 | 28 AA2: 29 | MFTFLHKARKFFPAFILLFTSFFFPNGV |
| | Probability: | |
| 13503, | 0.929 AA1: | |
| 13504 | 20 AA2: 21 | MFTGAFSGLAGMCFAYSIMA |
| | Probability: | |
| 37935, | 0.597 AA1: | |
| 37936 | 20 AA2: 21 | MFTMIESSAAWAESSSAALA |
| | Probability: | |
| 28803, | 0.795 AA1: | |
| 28804 | 27 AA2: 28 | MFTNLQISLLIPVSALIALFSALFFYK |
| | Probability: | |
| 47385, | 0.900 AA1: | |
| 47386 | 25 AA2: 26 | MFTRFHFVPLRAVVLFFLTSLCGYA |
| | Probability: | |
| 77149, | 1.000 AA1: | |
| 77150 | 21 AA2: 22 | MFTRNKWFVVGLAFAVALAFA |
| | Probability: | |
| 41551, | 0.999 AA1: | |
| 41552 | 20 AA2: 21 | MFTTIFFILFAAVCLVSAFA |
| | Probability: | |
| 12537, | 0.997 AA1: | |
| 12538 | 15 AA2: 16 | MFTVLITLAAAAATA |
| 57651, | Probability: | |
| 57652 | 0.984 AA1: | MFTYSTVMGTLAGAMSGASNE |

| | | |
|---|---|---|
| | 21 AA2: 22 | |
| | Probability: | |
| 16075, | 1.000 AA1: | |
| 16076 | 27 AA2: 28 | MFVVKFLMLILPAAVLLFSACTAKAQS |
| | Probability: | |
| 32427, | 0.949 AA1: | |
| 32428 | 26 AA2: 27 | MFYIFGLSGFLAIFHSSSMITRXLVA |
| | Probability: | |
| 40775, | 0.997 AA1: | |
| 40776 | 18 AA2: 19 | MGAAFAVAAAGFAGAAGA |
| | Probability: | |
| 36979, | 0.660 AA1: | |
| 36980 | 21 AA2: 22 | MGAAINIAVVICAAGASSRMK |
| | Probability: | |
| 23889, | 0.809 AA1: | |
| 23890 | 19 AA2: 20 | MGAHAVLLIAALLDAKALA |
| | Probability: | |
| 78357, | 0.639 AA1: | |
| 78358 | 22 AA2: 23 | MGALGVIFLGVSLLAMLLRVLA |
| | Probability: | |
| 33967, | 0.853 AA1: | |
| 33968 | 21 AA2: 22 | MGAPRALALPLGATGRRAALA |
| | Probability: | |
| 75161, | 0.590 AA1: | |
| 75162 | 25 AA2: 26 | MGCTCGFAATGCTGLGTGCAGGLTG |
| | Probability: | |
| 30655, | 1.000 AA1: | |
| 30656 | 28 AA2: 29 | MGDFMRRINRAILTAATFIAVAVGSAAA |
| | Probability: | |
| 28669, | 0.622 AA1: | |
| 28670 | 26 AA2: 27 | MGDILSQSEIDAMLSGLSASPAAAPA |
| | Probability: | |
| 40645, | 0.521 AA1: | |
| 40646 | 27 AA2: 28 | MGDLAGFAAAVFMRGIGFVQIPTTLLA |

| | | |
|---|---|---|
| | Probability: | |
| 62089, | 0.509 AA1: | |
| 62090 | 17 AA2: 18 | MGDTRLALSVPSLGTHA |
| | Probability: | |
| 17203, | 0.872 AA1: | |
| 17204 | 25 AA2: 26 | MGEILSQSEIDAMLSGLSTAPAATA |
| | Probability: | |
| 75963, | 0.907 AA1: | |
| 75964 | 23 AA2: 24 | MGEXISLLVSLSGVAFIALNLTG |
| | Probability: | |
| 45991, | 0.985 AA1: | |
| 45992 | 24 AA2: 25 | MGFFSDAVSLVWLFSGLLLLPVIA |
| | Probability: | |
| 73783, | 0.979 AA1: | |
| 73784 | 24 AA2: 25 | MGFLELLLIAVGLSMDAFAVSVSA |
| | Probability: | |
| 11327, | 0.998 AA1: | |
| 11328 | 19 AA2: 20 | MGFLFAAAISAFFPQGLYA |
| | Probability: | |
| 28571, | 0.537 AA1: | |
| 28572 | 28 AA2: 29 | MGFLIMNYKLVPFIVLTFIFVIALFFLQ |
| | Probability: | |
| 42113, | 1.000 AA1: | |
| 42114 | 26 AA2: 27 | MGFPKKVFCSFVIAVFVFCAASGLFA |
| | Probability: | |
| 1963, | 0.597 AA1: | |
| 1964 | 22 AA2: 23 | MGFSLLIAAVMFFIRIYQDASA |
| | Probability: | |
| 19743, | 0.787 AA1: | |
| 19744 | 26 AA2: 27 | MGFYMTVIFAAFSTAGQFFSLQMGFG |
| | Probability: | |
| 13913, | 0.726 AA1: | |
| 13914 | 28 AA2: 29 | MGGEKLKISIFALLLLTVGFLVQFQTIS |
| 4707, | Probability: | MGGGMRQAGILAAAGLLALTEHPSLLA |

| | | |
|---|---|---|
| 4708 | 0.564 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 10215, | 0.921 AA1: | |
| 10216 | 30 AA2: 31 | MGGIKKEYQKLNSNTTVLGSAVLAFASVPA |
| | Probability: | |
| 74931, | 0.989 AA1: | |
| 74932 | 23 AA2: 24 | MGGLGGAAVGTAGGFAAAGATTA |
| | Probability: | |
| 7793, | 0.980 AA1: | |
| 7794 | 28 AA2: 29 | MGGLSLVAPLMSLAMAFAMLFGIGSANM |
| | Probability: | |
| 43045, | 1.000 AA1: | |
| 43046 | 27 AA2: 28 | MGGRLTTAAFLGAAGFTAALSTAAAFG |
| | Probability: | |
| 13059, | 0.668 AA1: | |
| 13060 | 18 AA2: 19 | MGILAGILAGCGSQPVPG |
| | Probability: | |
| 68715, | 0.960 AA1: | |
| 68716 | 27 AA2: 28 | MGILSTVLLVFFVIIAILLILLVLVQN |
| | Probability: | |
| 30665, | 0.596 AA1: | |
| 30666 | 20 AA2: 21 | MGIQIASLYTGGSIPMPAQQ |
| | Probability: | |
| 5799, | 0.950 AA1: | |
| 5800 | 27 AA2: 28 | MGIRNGIVLIVVCVISIGFVCCTHLTA |
| | Probability: | |
| 77975, | 0.994 AA1: | |
| 77976 | 23 AA2: 24 | MGITPGLVLALALCLCAMGMARK |
| | Probability: | |
| 14595, | 0.807 AA1: | |
| 14596 | 24 AA2: 25 | MGIVLIILAFVLMIIGIIGAVVPA |
| 70855, | Probability: | |
| 70856 | 0.869 AA1: | MGIYIEALVLYILLFLHGSLALFAGGGAEA |

| | | |
|---|---|---|
| | 30 AA2: 31 | |
| | Probability: | |
| 59257, | 0.636 AA1: | |
| 59258 | 27 AA2: 28 | MGKAKFRDISRRNSLLFFLFLLPAVFA |
| | Probability: | |
| 60001, | 0.943 AA1: | |
| 60002 | 26 AA2: 27 | MGKDLKILKYLLCLLLVLPGLGRLEA |
| | Probability: | |
| 53423, | 0.999 AA1: | |
| 53424 | 23 AA2: 24 | MGKFARFGLTLLLTTATIVGAVA |
| | Probability: | |
| 71289, | 1.000 AA1: | |
| 71290 | 23 AA2: 24 | MGKIKAFAAVAALSLAVAGNLWA |
| | Probability: | |
| 69979, | 1.000 AA1: | |
| 69980 | 24 AA2: 25 | MGKKFLLTFFFFLFYFFFCISVYA |
| | Probability: | |
| 30891, | 0.803 AA1: | |
| 30892 | 28 AA2: 29 | MGKKLFAYLKLKPVLALIISILIFTGIA |
| | Probability: | |
| 77383, | 0.945 AA1: | |
| 77384 | 17 AA2: 18 | MGLAASACGFPPSSASE |
| | Probability: | |
| 11425, | 0.932 AA1: | |
| 11426 | 19 AA2: 20 | MGLALANSIAAVQAGCVHA |
| | Probability: | |
| 14559, | 0.950 AA1: | |
| 14560 | 19 AA2: 20 | MGLAPFSLSLMSSSSASLS |
| | Probability: | |
| 49985, | 0.942 AA1: | |
| 49986 | 26 AA2: 27 | MGLFAIIAAFAVLFVLVILILVKLYS |
| | Probability: | |
| 6709, | 0.988 AA1: | |
| 6710 | 31 AA2: 32 | MGLFFSLFPNPFTVLFCVSLLGVTLGFLAFN |

| | | |
|---|---|---|
| | Probability: | |
| 38229, | 0.949 AA1: | |
| 38230 | 25 AA2: 26 | MGLKKPKGLALGAATAPVEPAISLS |
| | Probability: | |
| 17971, | 0.774 AA1: | |
| 17972 | 20 AA2: 21 | MGLNAFFAFTVVLGMGYSWQ |
| | Probability: | |
| 41391, | 0.989 AA1: | |
| 41392 | 26 AA2: 27 | MGLSKKQKNLLSAVSLLFIFVFVIIA |
| | Probability: | |
| 74861, | 0.801 AA1: | |
| 74862 | 20 AA2: 21 | MGLVSVGLATAIGLFLGALA |
| | Probability: | |
| 39347, | 0.993 AA1: | |
| 39348 | 27 AA2: 28 | MGMGSGRSNSFLLFVLFVFAGLGSGWA |
| | Probability: | |
| 41757, | 0.948 AA1: | |
| 41758 | 24 AA2: 25 | MGMIGAYIVILVIVFFLGFFAIFA |
| | Probability: | |
| 1783, | 0.948 AA1: | |
| 1784 | 28 AA2: 29 | MGNRFLGRKIKKAFKIAGIVFIAVILMA |
| | Probability: | |
| 12095, | 0.732 AA1: | |
| 12096 | 16 AA2: 17 | MGNSATLIATVSPSNA |
| | Probability: | |
| 48675, | 0.996 AA1: | |
| 48676 | 21 AA2: 22 | MGNYTVILLIVFFLGFFAIFA |
| | Probability: | |
| 36975, | 0.796 AA1: | |
| 36976 | 20 AA2: 21 | MGPAAPFFGLLFGGGAGSRA |
| | Probability: | |
| 625, | 0.670 AA1: | |
| 626 | 24 AA2: 25 | MGQLASVAGGSALVLFTSYQSLQS |
| 32893, | Probability: | MGQNRHEIPYYVLSMVAMLLCLVVATTIMG |

| | | |
|---|---|---|
| 32894 | 0.755 AA1: | |
| | 30 AA2: 31 | |
| | Probability: | |
| 39843, | 0.962 AA1: | |
| 39844 | 29 AA2: 30 | MGQNSARDARARSRAAFFSVLAVAAVLAA |
| | Probability: | |
| 59187, | 0.792 AA1: | |
| 59188 | 24 AA2: 25 | MGREPFIKRFSALTIRGCIMKAAA |
| | Probability: | |
| 69913, | 1.000 AA1: | |
| 69914 | 24 AA2: 25 | MGRINRLGWALLLAAALAVMPVGA |
| | Probability: | |
| 24873, | 0.701 AA1: | |
| 24874 | 21 AA2: 22 | MGRKKVGAFFAQCLARATAWT |
| | Probability: | |
| 73407, | 1.000 AA1: | |
| 73408 | 19 AA2: 20 | MGRKVAAGAASPAPAPAGA |
| | Probability: | |
| 68693, | 0.674 AA1: | |
| 68694 | 30 AA2: 31 | MGRLYHKKKKSKRGLLVRLCAMLPPRSKHL |
| | Probability: | |
| 49181, | 0.741 AA1: | |
| 49182 | 20 AA2: 21 | MGRNSLSGLAFGLGTFISVG |
| | Probability: | |
| 30795, | 0.786 AA1: | |
| 30796 | 22 AA2: 23 | MGRTCSRRGVWALILIPSGVCA |
| | Probability: | |
| 15895, | 1.000 AA1: | |
| 15896 | 20 AA2: 21 | MGRVMTVLSAAAVLVGSAAA |
| | Probability: | |
| 11391, | 0.998 AA1: | |
| 11392 | 22 AA2: 23 | MGSKTIIXLXAVIFYFSGLLYA |
| 38797, | Probability: | |
| 38798 | 0.562 AA1: | MGSLFLFLLLANIIPVLSPLGLKAFG |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 50325, | 0.690 AA1: | |
| 50326 | 26 AA2: 27 | MGSLFLFLLLANIIPVLTPLGLKAFG |
| | Probability: | |
| 65159, | 0.551 AA1: | |
| 65160 | 16 AA2: 17 | MGSRASGGSFVGGVFA |
| | Probability: | |
| 37767, | 0.694 AA1: | |
| 37768 | 24 AA2: 25 | MGTKKVSFKGADAKFYLLSCPAHA |
| | Probability: | |
| 14553, | 0.999 AA1: | |
| 14554 | 22 AA2: 23 | MGTLTAALASSALVALVEAALA |
| | Probability: | |
| 11823, | 0.946 AA1: | |
| 11824 | 27 AA2: 28 | MGTRRSPRPTSSTATSAGASAPLRTSS |
| | Probability: | |
| 28617, | 0.991 AA1: | |
| 28618 | 25 AA2: 26 | MGVNVRRTRWLIICCSTLITASAVS |
| | Probability: | |
| 40959, | 0.998 AA1: | |
| 40960 | 26 AA2: 27 | MGVRSVMRKKASVLLAIVLASGCATN |
| | Probability: | |
| 1629, | 0.953 AA1: | |
| 1630 | 26 AA2: 27 | MGVTATATGGGAITAGSAGTTATSGA |
| | Probability: | |
| 30257, | 0.753 AA1: | |
| 30258 | 19 AA2: 20 | MGWLFWVWLVRLHAGSAGA |
| | Probability: | |
| 77455, | 0.692 AA1: | |
| 77456 | 20 AA2: 21 | MGXPRQIGTSIPAGAGSLMA |
| | Probability: | |
| 77363, | 0.967 AA1: | |
| 77364 | 23 AA2: 24 | MGXXQVKIGRRYFLFSAAAIALA |

| | | |
|---|---|---|
| | Probability: | |
| 40747, | 0.844 AA1: | |
| 40748 | 18 AA2: 19 | MGYLAGLITTSSMPNANS |
| | Probability: | |
| 16879, | 0.776 AA1: | |
| 16880 | 16 AA2: 17 | MGYLAGLVTTSGMSNA |
| | Probability: | |
| 30815, | 0.822 AA1: | |
| 30816 | 18 AA2: 19 | MGYRTIWALFGAANQLLA |
| | Probability: | |
| 60319, | 0.837 AA1: | |
| 60320 | 26 AA2: 27 | MGYYEINKKCAVCRLAVFWLSAVVLA |
| | Probability: | |
| 1001, | 0.993 AA1: | |
| 1002 | 32 AA2: 33 | MHAFAISVIRPPTFPACSTLWLSALSTSLVMA |
| | Probability: | |
| 48187, | 0.924 AA1: | |
| 48188 | 24 AA2: 25 | MHAQEAESLAATGAAVCLCPRSNA |
| | Probability: | |
| 45839, | 0.849 AA1: | |
| 45840 | 21 AA2: 22 | MHCHNDLGLAVANSIAAIANG |
| | Probability: | |
| 37157, | 0.842 AA1: | |
| 37158 | 20 AA2: 21 | MHDILVIGAGIAGLTVALKA |
| | Probability: | |
| 15405, | 0.899 AA1: | |
| 15406 | 21 AA2: 22 | MHFXYGAGSILSPLAAGYVAS |
| | Probability: | |
| 24653, | 0.995 AA1: | |
| 24654 | 30 AA2: 31 | MHHGREKRRFMKYPRLAAFALFALAAVSLA |
| | Probability: | |
| 77015, | 0.701 AA1: | |
| 77016 | 26 AA2: 27 | MHIAFHLAIDRIFSAVFTSTVLPALA |
| 30539, | Probability: | MHIMAAAAAARGSSGA |

| | | |
|---|---|---|
| 30540 | 0.983 AA1: | |
| | 16 AA2: 17 | |
| | Probability: | |
| 56859, | 0.775 AA1: | |
| 56860 | 18 AA2: 19 | MHKGHIWSALCVASVAKA |
| | Probability: | |
| 22031, | 0.536 AA1: | |
| 22032 | 21 AA2: 22 | MHLPLGAIIYFLVAGYTPLMA |
| | Probability: | |
| 78427, | 0.987 AA1: | |
| 78428 | 23 AA2: 24 | MHLSKAAALFGAVSLMLSNHCLT |
| | Probability: | |
| 61077, | 0.900 AA1: | |
| 61078 | 21 AA2: 22 | MHMGTAGAAWATVAGQTASFA |
| | Probability: | |
| 55657, | 0.867 AA1: | |
| 55658 | 27 AA2: 28 | MHNAKFIVPNAFTSLNFLLGAASICWS |
| | Probability: | |
| 15703, | 0.771 AA1: | |
| 15704 | 27 AA2: 28 | MHNAKFIVPNAFTSLNFLLGVASICWS |
| | Probability: | |
| 35331, | 0.775 AA1: | |
| 35332 | 23 AA2: 24 | MHNIKFSEPVVTAALTHALSVSA |
| | Probability: | |
| 70571, | 0.617 AA1: | |
| 70572 | 25 AA2: 26 | MHRSSSFMICMARILGAPDSVPAGS |
| | Probability: | |
| 19919, | 0.993 AA1: | |
| 19920 | 27 AA2: 28 | MHSSRFLTAVLLFFLCFSCTKFKPIQA |
| | Probability: | |
| 6871, | 0.918 AA1: | |
| 6872 | 23 AA2: 24 | MHVRDAARAFALALAADPAAVSG |
| 22737, | Probability: | |
| 22738 | 0.742 AA1: | MHXXXXXXXXXTLFASFCITGVGGFIAKA |

| | | |
|---|---|---|
| | 29 AA2: 30 | |
| | Probability: | |
| 9809, | 0.503 AA1: | |
| 9810 | 20 AA2: 21 | MIAIVAVIVMGIAIIQLNQA |
| | Probability: | |
| 54165, | 0.734 AA1: | |
| 54166 | 17 AA2: 18 | MIALHIVALFLMVLANA |
| | Probability: | |
| 11347, | 0.679 AA1: | |
| 11348 | 28 AA2: 29 | MIAPALTINNRFLAAFIAFTMNSAAYCA |
| | Probability: | |
| 72515, | 0.972 AA1: | |
| 72516 | 25 AA2: 26 | MIARGVLSAALILGGAVIGTGKIVA |
| | Probability: | |
| 4127, | 0.980 AA1: | |
| 4128 | 28 AA2: 29 | MIARLCASVSAVGAGLILSLFLALGKMS |
| | Probability: | |
| 34105, | 1.000 AA1: | |
| 34106 | 23 AA2: 24 | MIARPSLVLLGAIAFVLSALSFA |
| | Probability: | |
| 55125, | 0.997 AA1: | |
| 55126 | 31 AA2: 32 | MIARSAFAPLRDKMRVIALLCFMLIPRFVSA |
| | Probability: | |
| 77773, | 0.987 AA1: | |
| 77774 | 27 AA2: 28 | MIARWFWREWRSPSLLIVWLALSLAVA |
| | Probability: | |
| 29959, | 0.670 AA1: | |
| 29960 | 18 AA2: 19 | MIASVPSSIAFATSVSSA |
| | Probability: | |
| 60495, | 0.593 AA1: | |
| 60496 | 18 AA2: 19 | MIASVPSSTAFATSVSSA |
| | Probability: | |
| 4771, | 0.602 AA1: | |
| 4772 | 23 AA2: 24 | MIATVRSGYQTAGLGSGSAPLAA |

| | | |
|---|---|---|
| | Probability: | |
| 76039, | 0.689 AA1: | |
| 76040 | 27 AA2: 28 | MIAVDVLILGAGPAGLSAGIYAARAGA |
| | Probability: | |
| 56573, | 0.502 AA1: | |
| 56574 | 22 AA2: 23 | MIAXGRKPAAVVARGMGTPCVA |
| | Probability: | |
| 14991, | 0.809 AA1: | |
| 14992 | 27 AA2: 28 | MICFIPANKLTTSANVIMLQYSAPVWA |
| | Probability: | |
| 70741, | 0.993 AA1: | |
| 70742 | 24 AA2: 25 | MICKFFKLFLFSLFLSLSLSSLPA |
| | Probability: | |
| 33511, | 0.689 AA1: | |
| 33512 | 19 AA2: 20 | MICRLCSGSXLFLPESINA |
| | Probability: | |
| 25891, | 0.521 AA1: | |
| 25892 | 21 AA2: 22 | MICSLIFAASSLETPSFRGFG |
| | Probability: | |
| 45913, | 0.615 AA1: | |
| 45914 | 19 AA2: 20 | MIDCPPSLGILTLNGMAAA |
| | Probability: | |
| 76231, | 1.000 AA1: | |
| 76232 | 28 AA2: 29 | MIEMIPRYRFFLCLVFAFFSCFTTQVYS |
| | Probability: | |
| 67931, | 0.908 AA1: | |
| 67932 | 26 AA2: 27 | MIENAVLQNILSFGPVFLLIAVRAFA |
| | Probability: | |
| 9957, | 0.876 AA1: | |
| 9958 | 26 AA2: 27 | MIENQIINEILSSAPAFLLIAVRALA |
| | Probability: | |
| 8017, | 0.508 AA1: | |
| 8018 | 26 AA2: 27 | MIESQIINEILSSAPAFLLIAIRALA |
| 66283, | Probability: | MIFFLTIFMRKIFIFIIVLSVFSGCG |

| | | |
|---|---|---|
| 66284 | 0.795 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 27945, | 0.983 AA1: | |
| 27946 | 19 AA2: 20 | MIFGLGVSTYSGASGQALA |
| | Probability: | |
| 65069, | 0.998 AA1: | |
| 65070 | 21 AA2: 22 | MIFIKRPRILAALVFILLFSA |
| | Probability: | |
| 51159, | 0.910 AA1: | |
| 51160 | 23 AA2: 24 | MIFITALVTGILNVLFFPLGLLG |
| | Probability: | |
| 53463, | 0.999 AA1: | |
| 53464 | 21 AA2: 22 | MIFKKRPRILAAFAFVLLFSA |
| | Probability: | |
| 20043, | 0.989 AA1: | |
| 20044 | 26 AA2: 27 | MIFSGFAVGRSILLMATTSGTLAAFA |
| | Probability: | |
| 59047, | 0.859 AA1: | |
| 59048 | 24 AA2: 25 | MIFTFIYIVISLLALFLGIFQIVG |
| | Probability: | |
| 36697, | 0.823 AA1: | |
| 36698 | 24 AA2: 25 | MIFTLIYIVISLLALFLGIFQIVG |
| | Probability: | |
| 2023, | 0.991 AA1: | |
| 2024 | 24 AA2: 25 | MIFVLTVSSLSTLLLTTFVLTVSS |
| | Probability: | |
| 69197, | 0.656 AA1: | |
| 69198 | 29 AA2: 30 | MIFYYQEDNMKKICFMGLIFYAMTVNGYA |
| | Probability: | |
| 12101, | 0.969 AA1: | |
| 12102 | 21 AA2: 22 | MIGESCILVIVFFLGFFAIFA |
| 54747, | Probability: | |
| 54748 | 0.852 AA1: | MIGFVIAAFIAGIGGCVYA |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 14939, | 0.929 AA1: | |
| 14940 | 23 AA2: 24 | MIGGIHMTKHALLLIFCLVMLIA |
| | Probability: | |
| 73957, | 0.987 AA1: | |
| 73958 | 20 AA2: 21 | MIGIITFIVLLSMGSSSALV |
| | Probability: | |
| 27863, | 0.806 AA1: | |
| 27864 | 28 AA2: 29 | MIGKIYETMKKTIYTIITAALLFACIIA |
| | Probability: | |
| 8509, | 0.736 AA1: | |
| 8510 | 21 AA2: 22 | MIGNPGCYPTGGSLGAFPALA |
| | Probability: | |
| 4051, | 0.993 AA1: | |
| 4052 | 30 AA2: 31 | MIGQRKSVTQRTFITGFVLVALFAALISAG |
| | Probability: | |
| 51823, | 0.841 AA1: | |
| 51824 | 28 AA2: 29 | MIGVLVGISALISCAVIAVHSGMNKENA |
| | Probability: | |
| 12473, | 0.767 AA1: | |
| 12474 | 19 AA2: 20 | MIIGERPSVLPLLCALVLA |
| | Probability: | |
| 26163, | 0.972 AA1: | |
| 26164 | 19 AA2: 20 | MIIIFITAIFALAIAFVLG |
| | Probability: | |
| 53749, | 0.938 AA1: | |
| 53750 | 22 AA2: 23 | MIIKKALPIGAESFLFTFLVMA |
| | Probability: | |
| 53021, | 0.907 AA1: | |
| 53022 | 22 AA2: 23 | MIILHNNSMTAGIIGATGYAGA |
| | Probability: | |
| 9119, | 0.994 AA1: | |
| 9120 | 29 AA2: 30 | MIITKSCKKSKRRQCAAFISLFALHPQLS |

| | | |
|---|---|---|
| | Probability: | |
| 48823, | 0.781 AA1: | |
| 48824 | 15 AA2: 16 | MIIVTAVFLVFGVMA |
| | Probability: | |
| 65961, | 0.538 AA1: | |
| 65962 | 21 AA2: 22 | MIKGRLISVVCMAESSIFAFS |
| | Probability: | |
| 69133, | 0.727 AA1: | |
| 69134 | 26 AA2: 27 | MIKGRRNFAIFSIIACVLLWGFSFIS |
| | Probability: | |
| 3959, | 0.917 AA1: | |
| 3960 | 28 AA2: 29 | MIKGRRNFAIFSIIACVLLWGFSFISIK |
| | Probability: | |
| 31889, | 0.998 AA1: | |
| 31890 | 27 AA2: 28 | MIKIKKRKIIDGLIYGIMAAAMTAVLA |
| | Probability: | |
| 58285, | 1.000 AA1: | |
| 58286 | 25 AA2: 26 | MIKIKRAVFFFLLLTMTVFLVPAQT |
| | Probability: | |
| 49819, | 0.957 AA1: | |
| 49820 | 17 AA2: 18 | MIKKALLGALESGVIAA |
| | Probability: | |
| 1473, | 0.994 AA1: | |
| 1474 | 28 AA2: 29 | MIKKIGMPVKYKLFFLLLPFLILITVFA |
| | Probability: | |
| 56983, | 0.748 AA1: | |
| 56984 | 26 AA2: 27 | MIKKKLKYTVKFVFRRRVFVIMSLLA |
| | Probability: | |
| 2807, | 0.507 AA1: | |
| 2808 | 21 AA2: 22 | MIKKMREMAPMIMIVIIVAFV |
| | Probability: | |
| 22027, | 0.994 AA1: | |
| 22028 | 22 AA2: 23 | MIKKVILFTILAILFFCVEINA |
| 34669, | Probability: | MIKLTALFIAFFRVGAFT |

| | | |
|---|---|---|
| 34670 | 0.732 AA1: | |
| | 18 AA2: 19 | |
| | Probability: | |
| 19353, | 0.895 AA1: | |
| 19354 | 20 AA2: 21 | MIKNIIWTIILAVAAGILQS |
| | Probability: | |
| 67605, | 0.790 AA1: | |
| 67606 | 20 AA2: 21 | MIKNIIWTIILAVTAGILQS |
| | Probability: | |
| 67607, | 0.795 AA1: | |
| 67608 | 20 AA2: 21 | MIKNIIWTIILAVVAGILQS |
| | Probability: | |
| 22163, | 0.758 AA1: | |
| 22164 | 19 AA2: 20 | MIKNLLGKAVSAPFRSVMA |
| | Probability: | |
| 19111, | 0.997 AA1: | |
| 19112 | 21 AA2: 22 | MIKNSRIAQALLVTGASAVGA |
| | Probability: | |
| 72285, | 0.994 AA1: | |
| 72286 | 24 AA2: 25 | MIKPGKKILIIIFCITLFTNITMA |
| | Probability: | |
| 47345, | 0.991 AA1: | |
| 47346 | 23 AA2: 24 | MIKPLLFSFSCLFLITGLLPLNA |
| | Probability: | |
| 28331, | 0.891 AA1: | |
| 28332 | 27 AA2: 28 | MIKQKFLKPPWLEMIIMLAAVALSAVG |
| | Probability: | |
| 17471, | 1.000 AA1: | |
| 17472 | 24 AA2: 25 | MIKRWRKILLFIFCITLLSNITLA |
| | Probability: | |
| 7011, | 0.918 AA1: | |
| 7012 | 29 AA2: 30 | MIKTPRFFLIALSLMFGTSVFFIINPDLV |
| 29035, | Probability: | |
| 29036 | 0.830 AA1: | MIKTPRFYLIALSLMFGTSVFFIILPSLK |

| | | |
|---|---|---|
| | 29 AA2: 30 | |
| | Probability: | |
| 44923, | 0.994 AA1: | |
| 44924 | 26 AA2: 27 | MIKTSTVFHGKKTIALLGACCLFLSA |
| | Probability: | |
| 24801, | 0.983 AA1: | |
| 24802 | 18 AA2: 19 | MIKVRSTMGSLVPLSAFA |
| | Probability: | |
| 73151, | 0.533 AA1: | |
| 73152 | 23 AA2: 24 | MILAEGRLLNLGCATGHPSFVMS |
| | Probability: | |
| 53179, | 0.700 AA1: | |
| 53180 | 25 AA2: 26 | MILDVKNRYGLALSSVGAIGVSAMM |
| | Probability: | |
| 49581, | 0.789 AA1: | |
| 49582 | 20 AA2: 21 | MILFLYAFRSVLPVFSSILA |
| | Probability: | |
| 78301, | 0.561 AA1: | |
| 78302 | 17 AA2: 18 | MILIRFLLPLLLCNLPA |
| | Probability: | |
| 69721, | 0.980 AA1: | |
| 69722 | 24 AA2: 25 | MILKIRQLKIVLLLFALIFFSCKS |
| | Probability: | |
| 51379, | 1.000 AA1: | |
| 51380 | 24 AA2: 25 | MILKNNRRLLAIMAVIAVLITALA |
| | Probability: | |
| 30175, | 0.525 AA1: | |
| 30176 | 24 AA2: 25 | MILLSSRSRCWTALTIFVGLGTVS |
| | Probability: | |
| 69391, | 0.655 AA1: | |
| 69392 | 28 AA2: 29 | MILSSVSFFKTSALPKLAFVFTSSRASK |
| | Probability: | |
| 73371, | 0.836 AA1: | |
| 73372 | 19 AA2: 20 | MILTAICTLFSLAMTILCA |

| | | |
|---|---|---|
| | Probability: | |
| 36623, | 0.507 AA1: | |
| 36624 | 23 AA2: 24 | MILVLALCAAAVYGIVFFLKRSS |
| | Probability: | |
| 5295, | 0.540 AA1: | |
| 5296 | 20 AA2: 21 | MILYLAVVLVSLGIAIGFVA |
| | Probability: | |
| 28675, | 0.600 AA1: | |
| 28676 | 21 AA2: 22 | MIMAESYFFVGVAGVGMSALA |
| | Probability: | |
| 33753, | 0.543 AA1: | |
| 33754 | 23 AA2: 24 | MIMFLALNTSGLTLIPISIMVYR |
| | Probability: | |
| 77241, | 0.998 AA1: | |
| 77242 | 22 AA2: 23 | MIMFRHVHLLSGVIAFAAAAFA |
| | Probability: | |
| 55131, | 0.645 AA1: | |
| 55132 | 23 AA2: 24 | MIMIAAMGVGGGMVIGFARVLFS |
| | Probability: | |
| 41257, | 0.593 AA1: | |
| 41258 | 27 AA2: 28 | MIMIKRYLLLYFIVLTIPLCLGLLVWQ |
| | Probability: | |
| 675, | 0.536 AA1: | |
| 676 | 27 AA2: 28 | MIMITSDFLIRIPAILLALTLHELSHA |
| | Probability: | |
| 55323, | 0.669 AA1: | |
| 55324 | 26 AA2: 27 | MIMKTTAKKILLTGGGTAGHVTPNLA |
| | Probability: | |
| 62861, | 0.848 AA1: | |
| 62862 | 18 AA2: 19 | MIMXGAMAGIGGALLFLA |
| | Probability: | |
| 44473, | 1.000 AA1: | |
| 44474 | 24 AA2: 25 | MINCKMTFFLAVFMLFSPLFSLWA |
| 44205, | Probability: | MINDMTVISLCAASYGYG |

| | | |
|---|---|---|
| 44206 | 0.530 AA1: | |
| | 18 AA2: 19 | |
| | Probability: | |
| 32547, | 0.839 AA1: | |
| 32548 | 22 AA2: 23 | MINFKFSDSSVNALSSVTAVFA |
| | Probability: | |
| 68947, | 0.759 AA1: | |
| 68948 | 21 AA2: 22 | MINKKLIIIPLLCILFIGCVK |
| | Probability: | |
| 72211, | 0.733 AA1: | |
| 72212 | 19 AA2: 20 | MINKRKTKIVCSLGPASSS |
| | Probability: | |
| 60493, | 0.858 AA1: | |
| 60494 | 23 AA2: 24 | MINKRKTKIVCSLGPASSSDSVV |
| | Probability: | |
| 3187, | 0.828 AA1: | |
| 3188 | 22 AA2: 23 | MINPLFSNVVVAISGSDASVLA |
| | Probability: | |
| 15103, | 0.999 AA1: | |
| 15104 | 23 AA2: 24 | MINQRKFITGLVLTALFAALISA |
| | Probability: | |
| 11795, | 0.705 AA1: | |
| 11796 | 22 AA2: 23 | MINSRNTATVSGRTSIGGIAGA |
| | Probability: | |
| 5353, | 0.513 AA1: | |
| 5354 | 28 AA2: 29 | MIPAFTMSTHSILAASKPIPFFSFFALS |
| | Probability: | |
| 26079, | 0.687 AA1: | |
| 26080 | 22 AA2: 23 | MIPFEQAKKIRASILFAGPLLA |
| | Probability: | |
| 75539, | 0.999 AA1: | |
| 75540 | 26 AA2: 27 | MIPMNTVSQSLRAICLLAIFSVLGMA |
| 60511, | Probability: | |
| 60512 | 0.812 AA1: | MIPRRLIDTAFTWLFRAFSFTGVA |

| | | |
|---|---|---|
| | 24 AA2: 25 | |
| | Probability: | |
| 21197, | 0.661 AA1: | |
| 21198 | 24 AA2: 25 | MIRCVIWTVILCISAGILQSTLIA |
| | Probability: | |
| 35213, | 0.992 AA1: | |
| 35214 | 18 AA2: 19 | MIRFILMVLGTALAAVLS |
| | Probability: | |
| 57513, | 1.000 AA1: | |
| 57514 | 28 AA2: 29 | MIRHTYNMVKKISFLFLLFVIVAGFAFS |
| | Probability: | |
| 54635, | 0.510 AA1: | |
| 54636 | 17 AA2: 18 | MIRKRLLAGLFSLLPLG |
| | Probability: | |
| 22167, | 0.663 AA1: | |
| 22168 | 20 AA2: 21 | MIRNMQATGTLSLVSSYTVN |
| | Probability: | |
| 347, | 0.997 AA1: | |
| 348 | 23 AA2: 24 | MIRQMGPLLTAIILAGRTGSATT |
| | Probability: | |
| 59087, | 1.000 AA1: | |
| 59088 | 20 AA2: 21 | MIRRLSSFCLFFSLFSSLYA |
| | Probability: | |
| 18845, | 1.000 AA1: | |
| 18846 | 21 AA2: 22 | MIRRVLLSFCLFFPLFFPVYA |
| | Probability: | |
| 45619, | 0.951 AA1: | |
| 45620 | 18 AA2: 19 | MIRSAEALPCMTIAAAFA |
| | Probability: | |
| 4367, | 0.999 AA1: | |
| 4368 | 27 AA2: 28 | MIRSAECTALAASILGSLLLSSFMAFS |
| | Probability: | |
| 35293, | 0.898 AA1: | |
| 35294 | 23 AA2: 24 | MIRSAEEIAIAISAVVLALLILA |

| | | |
|---|---|---|
| | Probability: | |
| 21077, | 0.987 AA1: | |
| 21078 | 29 AA2: 30 | MIRSAELRAFLLLFSAGFAVMSGFQMIIA |
| | Probability: | |
| 35943, | 0.925 AA1: | |
| 35944 | 25 AA2: 26 | MIRSAEPTGVAANSFTVAGATATNA |
| | Probability: | |
| 76059, | 0.673 AA1: | |
| 76060 | 25 AA2: 26 | MIRSAESASAALANVNAGCGSIVVS |
| | Probability: | |
| 77001, | 0.510 AA1: | |
| 77002 | 20 AA2: 21 | MIRSAESIALGLLPLALSVG |
| | Probability: | |
| 37533, | 0.947 AA1: | |
| 37534 | 22 AA2: 23 | MIRSAEVLAFALMVALSSYFMT |
| | Probability: | |
| 38431, | 1.000 AA1: | |
| 38432 | 24 AA2: 25 | MIRSARVRFIGMLASASLGAPAQG |
| | Probability: | |
| 10873, | 0.987 AA1: | |
| 10874 | 24 AA2: 25 | MIRSIIWTVIICVAFAILQSTLLA |
| | Probability: | |
| 50665, | 1.000 AA1: | |
| 50666 | 23 AA2: 24 | MIRSLLLIALMAVLSFAAAPASA |
| | Probability: | |
| 13181, | 0.901 AA1: | |
| 13182 | 28 AA2: 29 | MIRVAAYVSIAMLIAVIIVAVIMSLSIS |
| | Probability: | |
| 62269, | 0.647 AA1: | |
| 62270 | 24 AA2: 25 | MIRWPRGMTPSSWICCSMSTXRFN |
| | Probability: | |
| 12117, | 0.522 AA1: | |
| 12118 | 20 AA2: 21 | MIRWPSCRQFMFISSASAYQ |
| 10063, | Probability: | MISGSAVLGLFMGVILWFATPSIV |

| | | |
|---|---|---|
| 10064 | 0.945 AA1: | |
| | 24 AA2: 25 | |
| | Probability: | |
| 33705, | 0.991 AA1: | |
| 33706 | 16 AA2: 17 | MISIIILSFMAGFAFA |
| | Probability: | |
| 35947, | 0.996 AA1: | |
| 35948 | 26 AA2: 27 | MISKLTKFVQMAAIGFALTLTLSCSG |
| | Probability: | |
| 16343, | 0.986 AA1: | |
| 16344 | 18 AA2: 19 | MISLPAVIAGLTAGLSCA |
| | Probability: | |
| 40407, | 1.000 AA1: | |
| 40408 | 26 AA2: 27 | MISLVKKPLVFAAFALVLLFSACASS |
| | Probability: | |
| 24827, | 0.965 AA1: | |
| 24828 | 21 AA2: 22 | MISMARTLGAPLTVPAGKVAA |
| | Probability: | |
| 1209, | 0.817 AA1: | |
| 1210 | 29 AA2: 30 | MISRIFIARPRFAVVVSLIITIAGLIALL |
| | Probability: | |
| 31253, | 1.000 AA1: | |
| 31254 | 20 AA2: 21 | MISRLFFAVFLTILSAITSA |
| | Probability: | |
| 6049, | 0.995 AA1: | |
| 6050 | 21 AA2: 22 | MISSAPPGSIVMYATSANSVA |
| | Probability: | |
| 74769, | 0.967 AA1: | |
| 74770 | 24 AA2: 25 | MISSSSEAGVSTSSSIVISSSSEA |
| | Probability: | |
| 55897, | 0.743 AA1: | |
| 55898 | 18 AA2: 19 | MISTMLLIIIYLSFISLG |
| 53673, | Probability: | |
| 53674 | 0.805 AA1: | MITFFQIFLTLPVVFSLFSATRKFA |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 53979, | 0.687 AA1: | |
| 53980 | 19 AA2: 20 | MITFVFLSFFSPCRHSTSS |
| | Probability: | |
| 14787, | 0.820 AA1: | |
| 14788 | 19 AA2: 20 | MITFVFLSFLSPCRHSTSS |
| | Probability: | |
| 55797, | 0.660 AA1: | |
| 55798 | 18 AA2: 19 | MITGIIGLGLIGGSFAKA |
| | Probability: | |
| 51665, | 0.756 AA1: | |
| 51666 | 28 AA2: 29 | MITGIIGLGLIGGSFAKAYKAAGHTVFA |
| | Probability: | |
| 50277, | 0.549 AA1: | |
| 50278 | 19 AA2: 20 | MITSTSACIAMRARSSLYG |
| | Probability: | |
| 10781, | 1.000 AA1: | |
| 10782 | 20 AA2: 21 | MITVAAFCAACLGCIASAQQ |
| | Probability: | |
| 34183, | 0.988 AA1: | |
| 34184 | 23 AA2: 24 | MIVAKPRFLVLLFFCGLSVLIFA |
| | Probability: | |
| 74385, | 0.961 AA1: | |
| 74386 | 24 AA2: 25 | MIVFLQGIITGVMAAVFFAAGTVA |
| | Probability: | |
| 59337, | 0.658 AA1: | |
| 59338 | 18 AA2: 19 | MIVGVAGLGLIGGSLAKA |
| | Probability: | |
| 60787, | 0.998 AA1: | |
| 60788 | 26 AA2: 27 | MIVQYHKIAFLAGFALAMSFIFSCSS |
| | Probability: | |
| 77561, | 0.675 AA1: | |
| 77562 | 20 AA2: 21 | MIYGRAPLAFAFDFLAIALA |

| | | |
|---|---|---|
| | Probability: | |
| 34347, | 0.928 AA1: | |
| 34348 | 29 AA2: 30 | MIYITPYPKGSTTMRKLILTTAAVAAAIA |
| | Probability: | |
| 39231, | 0.990 AA1: | |
| 39232 | 18 AA2: 19 | MIYRKMALFCILPLLLQA |
| | Probability: | |
| 24145, | 0.887 AA1: | |
| 24146 | 21 AA2: 22 | MIYSEKTCAAAAVFTTNRVKA |
| | Probability: | |
| 59265, | 0.998 AA1: | |
| 59266 | 28 AA2: 29 | MIYSLRTKILAIGLAFFALIGTAFVLYS |
| | Probability: | |
| 57449, | 0.791 AA1: | |
| 57450 | 28 AA2: 29 | MKAGVICAGAAAGIAVSLVPIIVTLIVA |
| | Probability: | |
| 9437, | 0.790 AA1: | |
| 9438 | 20 AA2: 21 | MKAINLFPFLFTFLILACSN |
| | Probability: | |
| 12689, | 1.000 AA1: | |
| 12690 | 23 AA2: 24 | MKAKHFLVFVLALIATFTLGCAA |
| | Probability: | |
| 76379, | 0.617 AA1: | |
| 76380 | 23 AA2: 24 | MKALAAATGPVGLGILAVGTACV |
| | Probability: | |
| 67609, | 0.598 AA1: | |
| 67610 | 20 AA2: 21 | MKALFIGGTGTISSAISRLA |
| | Probability: | |
| 51909, | 0.880 AA1: | |
| 51910 | 23 AA2: 24 | MKATKTIMPLSALFFALSLYLLL |
| | Probability: | |
| 53241, | 0.549 AA1: | |
| 53242 | 26 AA2: 27 | MKAVKKIINDIMFVTGAALVILAVFS |
| 10695, | Probability: | MKCDVILAGVGGQGVLSVAAIIAQA |

| | | |
|---|---|---|
| 10696 | 0.681 AA1: 25 AA2: 26 Probability: | |
| 65205, 65206 | 0.559 AA1: 25 AA2: 26 Probability: | MKCDVILAGVGGQGVLSVAAIIAQA |
| 13041, 13042 | 1.000 AA1: 26 AA2: 27 Probability: | MKDIIKFRFTGILIMLAAIAAFSAAA |
| 36067, 36068 | 1.000 AA1: 28 AA2: 29 Probability: | MKDKKIKKIIQLFTAIFIFVMSSAPLFA |
| 32785, 32786 | 0.724 AA1: 20 AA2: 21 Probability: | MKDTLPAVALCATGAEKAVS |
| 49307, 49308 | 0.991 AA1: 23 AA2: 24 Probability: | MKDVIKFRFAGIGIVLAMCAVFA |
| 35721, 35722 | 0.987 AA1: 29 AA2: 30 Probability: | MKDVLTNIGRRRKMKKIIFILLLTGLCAA |
| 59623, 59624 | 0.572 AA1: 17 AA2: 18 Probability: | MKDXILLCSWLISCVLK |
| 71327, 71328 | 0.884 AA1: 26 AA2: 27 Probability: | MKEAKKMKNKKLFAKFAMIAVILITA |
| 44081, 44082 | 1.000 AA1: 26 AA2: 27 Probability: | MKEIIKFRFTGILIMLAAIAAFSAAA |
| 61427, 61428 | 0.583 AA1: 14 AA2: 15 Probability: | MKEKKIFLILGSHA |
| 67239, 67240 | 0.815 AA1: | MKEKKMSRKTTFKNAVVVILAALFLLM |

| | | |
|---|---|---|
| | 27 AA2: 28 | |
| | Probability: | |
| 69103, | 0.870 AA1: | |
| 69104 | 21 AA2: 22 | MKELNFLRAALIAGITVVSIA |
| | Probability: | |
| 28295, | 0.657 AA1: | |
| 28296 | 21 AA2: 22 | MKELQYDAAVIGGGAAGMAAA |
| | Probability: | |
| 69973, | 0.697 AA1: | |
| 69974 | 21 AA2: 22 | MKELQYDAAVIGGGAAGMAAA |
| | Probability: | |
| 47237, | 0.997 AA1: | |
| 47238 | 23 AA2: 24 | MKEPIKSRVTGILIMLAAAAVFG |
| | Probability: | |
| 33627, | 0.998 AA1: | |
| 33628 | 23 AA2: 24 | MKETIKSRFPGILIMLAAAAVFG |
| | Probability: | |
| 59049, | 0.999 AA1: | |
| 59050 | 23 AA2: 24 | MKETIKSRFTGILIMLAAAAVFG |
| | Probability: | |
| 49325, | 0.985 AA1: | |
| 49326 | 25 AA2: 26 | MKEYKMKIGFKLTVIMVALGLFAIA |
| | Probability: | |
| 1081, | 1.000 AA1: | |
| 1082 | 23 AA2: 24 | MKEYTLFFACCLVSLIAAATVSA |
| | Probability: | |
| 77235, | 0.507 AA1: | |
| 77236 | 17 AA2: 18 | MKFAIIGXAVFMIVASX |
| | Probability: | |
| 19233, | 0.865 AA1: | |
| 19234 | 26 AA2: 27 | MKFAKYLFSLWAGVLIYASLFLITGA |
| | Probability: | |
| 50475, | 0.669 AA1: | |
| 50476 | 21 AA2: 22 | MKFATERSFIPSYALTFTVFA |

| | | |
|---|---|---|
| | Probability: | |
| 30413, | 0.628 AA1: | |
| 30414 | 17 AA2: 18 | MKFKFLLVLTVFFYCEA |
| | Probability: | |
| 60307, | 0.958 AA1: | |
| 60308 | 23 AA2: 24 | MKFKIKLSIIMIAIVAVIVGGIA |
| | Probability: | |
| 24457, | 0.999 AA1: | |
| 24458 | 18 AA2: 19 | MKFKIKMSCMMIAIVAIA |
| | Probability: | |
| 14393, | 1.000 AA1: | |
| 14394 | 22 AA2: 23 | MKFKKITIFLAVVFWAAGNVGA |
| | Probability: | |
| 59621, | 0.999 AA1: | |
| 59622 | 24 AA2: 25 | MKFKLKSNILMTAIMTLMVTGVAA |
| | Probability: | |
| 66617, | 1.000 AA1: | |
| 66618 | 23 AA2: 24 | MKFLKSFSLLIVLVGVLNSGALA |
| | Probability: | |
| 72373, | 1.000 AA1: | |
| 72374 | 23 AA2: 24 | MKFLKSFSLLIVLVGVLNSGALA |
| | Probability: | |
| 31533, | 0.838 AA1: | |
| 31534 | 24 AA2: 25 | MKFLRRHWYNIGLVIAIITGAYLA |
| | Probability: | |
| 51359, | 1.000 AA1: | |
| 51360 | 25 AA2: 26 | MKFMKYKILGFFIIIFLFVVTSAFS |
| | Probability: | |
| 19983, | 0.738 AA1: | |
| 19984 | 25 AA2: 26 | MKFNIRRFADQIVGVFIIISFVALA |
| | Probability: | |
| 67853, | 1.000 AA1: | |
| 67854 | 24 AA2: 25 | MKFNRLHILCLISLFTFVCFNIYA |
| 65777, | Probability: | MKFRVIITILIFFFLIISFLSA |

| | | |
|---|---|---|
| 65778 | 1.000 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 22331, | 0.998 AA1: | |
| 22332 | 27 AA2: 28 | MKGGHCRVLLVALFAIMVLAAGPIVHA |
| | Probability: | |
| 29891, | 0.969 AA1: | |
| 29892 | 21 AA2: 22 | MKGILLAGGAGSRLYPLSAVA |
| | Probability: | |
| 56649, | 0.969 AA1: | |
| 56650 | 21 AA2: 22 | MKGILLAGGAGSRLYPLSAVA |
| | Probability: | |
| 23287, | 0.644 AA1: | |
| 23288 | 22 AA2: 23 | MKGILPPTGRTFTSVAMGTVSA |
| | Probability: | |
| 77403, | 0.962 AA1: | |
| 77404 | 26 AA2: 27 | MKGKFKVLLLVSVTILMLIDYTEVFT |
| | Probability: | |
| 66505, | 0.993 AA1: | |
| 66506 | 25 AA2: 26 | MKGLRTITAFIAGIVLFSAVTPLNA |
| | Probability: | |
| 44567, | 0.511 AA1: | |
| 44568 | 18 AA2: 19 | MKGPILVVLAAGMASRYG |
| | Probability: | |
| 44107, | 0.997 AA1: | |
| 44108 | 20 AA2: 21 | MKGRFIIPVIFFLIISFLHA |
| | Probability: | |
| 32815, | 0.995 AA1: | |
| 32816 | 24 AA2: 25 | MKGRRWDALFLAPVAGRAAAPVLA |
| | Probability: | |
| 24891, | 0.998 AA1: | |
| 24892 | 24 AA2: 25 | MKHFPACFLFLPLAFLFLSCASVG |
| 76767, | Probability: | |
| 76768 | 1.000 AA1: | MKHFSKFFAIILLASLAGNASA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 2827, | 0.685 AA1: | |
| 2828 | 14 AA2: 15 | MKHIIIGAGAAAMS |
| | Probability: | |
| 11083, | 0.979 AA1: | |
| 11084 | 20 AA2: 21 | MKHLFAFFAVVVICGCAALT |
| | Probability: | |
| 9751, | 0.950 AA1: | |
| 9752 | 26 AA2: 27 | MKHNKIFTLIAATAIILGILNSYLIM |
| | Probability: | |
| 46997, | 0.998 AA1: | |
| 46998 | 23 AA2: 24 | MKHNKSTSFATAVIVSILFTACA |
| | Probability: | |
| 6275, | 0.990 AA1: | |
| 6276 | 22 AA2: 23 | MKHTFWIIALTVLLCIFAGCKA |
| | Probability: | |
| 69315, | 0.979 AA1: | |
| 69316 | 23 AA2: 24 | MKIAKKALAALTVITIVALIGCA |
| | Probability: | |
| 70435, | 0.979 AA1: | |
| 70436 | 23 AA2: 24 | MKIAKKALAALTVITIVALIGCA |
| | Probability: | |
| 56603, | 0.995 AA1: | |
| 56604 | 16 AA2: 17 | MKIFYLIFLLVLSAHA |
| | Probability: | |
| 56169, | 0.998 AA1: | |
| 56170 | 23 AA2: 24 | MKIGFKLTALMTALGIFAVASVS |
| | Probability: | |
| 58535, | 0.998 AA1: | |
| 58536 | 23 AA2: 24 | MKIGFKLTALMTTLGIFAVASVS |
| | Probability: | |
| 59493, | 0.845 AA1: | |
| 59494 | 16 AA2: 17 | MKIGIIGLGLIGGSFA |

| | | |
|---|---|---|
| | Probability: | |
| 74993, | 0.999 AA1: | |
| 74994 | 17 AA2: 18 | MKIGSIAASAASVIAMA |
| | Probability: | |
| 1579, | 0.974 AA1: | |
| 1580 | 20 AA2: 21 | MKIGVMGAALSTFLSRIIAA |
| | Probability: | |
| 12991, | 0.989 AA1: | |
| 12992 | 25 AA2: 26 | MKIIIPNSRIFLFSLFFLSINILFG |
| | Probability: | |
| 1323, | 0.905 AA1: | |
| 1324 | 20 AA2: 21 | MKIKEIPPVSAASSAQAPSA |
| | Probability: | |
| 52341, | 1.000 AA1: | |
| 52342 | 25 AA2: 26 | MKIKFAPLFLLLLLCSCLTINRSRA |
| | Probability: | |
| 22119, | 1.000 AA1: | |
| 22120 | 24 AA2: 25 | MKIKFINLVAGSLLAMVFTFSCSG |
| | Probability: | |
| 57307, | 0.989 AA1: | |
| 57308 | 23 AA2: 24 | MKIKFKLSILIIAIVVVIVTGVA |
| | Probability: | |
| 14695, | 0.999 AA1: | |
| 14696 | 20 AA2: 21 | MKIKFKLSIVVIAIMAVVVA |
| | Probability: | |
| 65481, | 0.998 AA1: | |
| 65482 | 23 AA2: 24 | MKIKIKLSIMMIAIVAIVAGGVA |
| | Probability: | |
| 62627, | 0.991 AA1: | |
| 62628 | 27 AA2: 28 | MKIKIKLSIMMIAIVLIVSGGLTVLQM |
| | Probability: | |
| 66693, | 0.999 AA1: | |
| 66694 | 19 AA2: 20 | MKIKIKLSLMMIAIVAVVA |
| 32801, | Probability: | MKIKLKSIVIIAVIIFFLVLIGPFYV |

| | | |
|---|---|---|
| 32802 | 0.947 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 54051, | 0.997 AA1: | |
| 54052 | 19 AA2: 20 | MKIKNLFIIGLLFFCFACA |
| | Probability: | |
| 34411, | 1.000 AA1: | |
| 34412 | 23 AA2: 24 | MKIKRNIILFSFLAGLLSATVLG |
| | Probability: | |
| 75711, | 1.000 AA1: | |
| 75712 | 18 AA2: 19 | MKIKTMLILAVFTGAVFS |
| | Probability: | |
| 8107, | 0.998 AA1: | |
| 8108 | 22 AA2: 23 | MKIKYLALLLLILLTTCNGTEG |
| | Probability: | |
| 35075, | 0.995 AA1: | |
| 35076 | 22 AA2: 23 | MKIKYNAPTVLTYTLLCALVMA |
| | Probability: | |
| 37127, | 1.000 AA1: | |
| 37128 | 23 AA2: 24 | MKILGIFLLLFSFSIIFSPAVYS |
| | Probability: | |
| 42453, | 0.915 AA1: | |
| 42454 | 24 AA2: 25 | MKILKYVIGLLLAAIGLYIFFRGS |
| | Probability: | |
| 62855, | 0.772 AA1: | |
| 62856 | 23 AA2: 24 | MKINLQKGITLLNSGLTPSVINA |
| | Probability: | |
| 55059, | 0.980 AA1: | |
| 55060 | 28 AA2: 29 | MKINLQKGITLLNSGLTPSVINALSSVA |
| | Probability: | |
| 56779, | 0.989 AA1: | |
| 56780 | 28 AA2: 29 | MKINRCARNRFVFAAVFAALIAVSNFFQ |
| 65621, | Probability: | |
| 65622 | 0.955 AA1: | MKINRISKRLPLILCSVLLAFLIIT |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 50255, | 1.000 AA1: | |
| 50256 | 17 AA2: 18 | MKIQALVFLAVAASLFA |
| | Probability: | |
| 47067, | 0.999 AA1: | |
| 47068 | 22 AA2: 23 | MKIRFKIILVVLPVVILALCIA |
| | Probability: | |
| 37139, | 0.999 AA1: | |
| 37140 | 22 AA2: 23 | MKIRFKIILVVLPVVVLALCIA |
| | Probability: | |
| 72829, | 0.996 AA1: | |
| 72830 | 24 AA2: 25 | MKIRHFIPHVIVFLMFVCVCADLA |
| | Probability: | |
| 7117, | 1.000 AA1: | |
| 7118 | 30 AA2: 31 | MKIRIKLSLMVIAIVSVVAGGIAIILLLQA |
| | Probability: | |
| 2199, | 0.987 AA1: | |
| 2200 | 26 AA2: 27 | MKIRLTLIIIASLLLFQGLSPWEGSA |
| | Probability: | |
| 48787, | 0.997 AA1: | |
| 48788 | 25 AA2: 26 | MKIRRKACAKGIPCLTSIDTANALA |
| | Probability: | |
| 71225, | 0.998 AA1: | |
| 71226 | 20 AA2: 21 | MKIRVKLTAIMVALGLFAIA |
| | Probability: | |
| 62249, | 1.000 AA1: | |
| 62250 | 22 AA2: 23 | MKIRVKMCLALLPLIIVTLALA |
| | Probability: | |
| 23881, | 0.969 AA1: | |
| 23882 | 25 AA2: 26 | MKISAYYLPKSLSYALPISLLFASA |
| | Probability: | |
| 50965, | 0.999 AA1: | |
| 50966 | 20 AA2: 21 | MKISFKLSVILVALGLFAIA |

| | | |
|---|---|---|
| | Probability: | |
| 72709, | 0.785 AA1: | |
| 72710 | 23 AA2: 24 | MKITDEAGIMAGQAAAFAGNTTA |
| | Probability: | |
| 41451, | 0.780 AA1: | |
| 41452 | 24 AA2: 25 | MKITTAQNYFTFFAFLTGAFFTGA |
| | Probability: | |
| 53671, | 0.723 AA1: | |
| 53672 | 17 AA2: 18 | MKIVLLLFALIFFSCKS |
| | Probability: | |
| 30377, | 0.999 AA1: | |
| 30378 | 18 AA2: 19 | MKIVLVGFGLLASSVASA |
| | Probability: | |
| 53227, | 0.995 AA1: | |
| 53228 | 20 AA2: 21 | MKKAFFFLFLIFSFYGVCFS |
| | Probability: | |
| 69589, | 0.998 AA1: | |
| 69590 | 20 AA2: 21 | MKKAFFFLLLIFSFYGVCFS |
| | Probability: | |
| 66711, | 0.985 AA1: | |
| 66712 | 26 AA2: 27 | MKKAFFYFSVGIAALSVLLIIFLATG |
| | Probability: | |
| 15931, | 0.856 AA1: | |
| 15932 | 16 AA2: 17 | MKKAGGFFLVLGGSQG |
| | Probability: | |
| 12555, | 0.998 AA1: | |
| 12556 | 24 AA2: 25 | MKKAIFAAIVLTFTVSCTAPSSMV |
| | Probability: | |
| 2055, | 1.000 AA1: | |
| 2056 | 19 AA2: 20 | MKKAIIAFVFILAGVSAFA |
| | Probability: | |
| 71563, | 0.954 AA1: | |
| 71564 | 21 AA2: 22 | MKKAIIGLLIVSLFFTAGCTK |
| 36301, | Probability: | MKKAIXAVIVLVVIVLVFVLYA |

| | | |
|---|---|---|
| 36302 | 1.000 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 53721, | 1.000 AA1: | |
| 53722 | 28 AA2: 29 | MKKALFLLVLLFLFPVFAILPQSAGTAA |
| | Probability: | |
| 4615, | 1.000 AA1: | |
| 4616 | 19 AA2: 20 | MKKALIVLFALFVCTGLFA |
| | Probability: | |
| 22879, | 0.610 AA1: | |
| 22880 | 17 AA2: 18 | MKKDLGLIAMANSNAYV |
| | Probability: | |
| 53545, | 0.866 AA1: | |
| 53546 | 18 AA2: 19 | MKKEILTASLVDVLSSVA |
| | Probability: | |
| 5051, | 0.793 AA1: | |
| 5052 | 30 AA2: 31 | MKKEKIKNKKGRFSSSLSRILISFFFFLFS |
| | Probability: | |
| 73603, | 0.993 AA1: | |
| 73604 | 24 AA2: 25 | MKKFFFTLFLLLVIGGVIFMIGWA |
| | Probability: | |
| 39131, | 0.724 AA1: | |
| 39132 | 28 AA2: 29 | MKKFFNYIKARPVAMVSLVCLAILYLMM |
| | Probability: | |
| 44099, | 0.987 AA1: | |
| 44100 | 19 AA2: 20 | MKKFIILPLLFLSLVSCEA |
| | Probability: | |
| 38837, | 1.000 AA1: | |
| 38838 | 20 AA2: 21 | MKKFITAVIVLFVCAVTGFA |
| | Probability: | |
| 25917, | 1.000 AA1: | |
| 25918 | 20 AA2: 21 | MKKFITVVIVLFVCAGAGFA |
| 20451, | Probability: | |
| 20452 | 0.628 AA1: | MKKFLITIFVTVFSIISVK |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 45851, | 0.988 AA1: | |
| 45852 | 20 AA2: 21 | MKKFLTICLLAFLAAGMLFV |
| | Probability: | |
| 42259, | 1.000 AA1: | |
| 42260 | 23 AA2: 24 | MKKFNLIIIAVLFFLLISCAGSA |
| | Probability: | |
| 56397, | 0.582 AA1: | |
| 56398 | 28 AA2: 29 | MKKFVTTIVVIFVIFIGLMIASPFYIIN |
| | Probability: | |
| 76245, | 1.000 AA1: | |
| 76246 | 21 AA2: 22 | MKKGFFLVAAVFAAFALGACA |
| | Probability: | |
| 34015, | 0.949 AA1: | |
| 34016 | 24 AA2: 25 | MKKGITLFTLVFFSLFLITNEVFS |
| | Probability: | |
| 29063, | 0.936 AA1: | |
| 29064 | 19 AA2: 20 | MKKGKFRLIILISIFAVFT |
| | Probability: | |
| 60745, | 0.976 AA1: | |
| 60746 | 23 AA2: 24 | MKKGSIFIFILYTFLYVSGSVHG |
| | Probability: | |
| 58297, | 1.000 AA1: | |
| 58298 | 27 AA2: 28 | MKKHNTSLLLLFLSVFFLICACSNGYK |
| | Probability: | |
| 38709, | 1.000 AA1: | |
| 38710 | 17 AA2: 18 | MKKIAPLVLFVVFTAFA |
| | Probability: | |
| 501, | 0.958 AA1: | |
| 502 | 24 AA2: 25 | MKKIDILIICIILFLLFGCTFDYG |
| | Probability: | |
| 65115, | 1.000 AA1: | |
| 65116 | 20 AA2: 21 | MKKIFCAFILLLFISVFAFS |

| | | |
|---|---|---|
| | Probability: | |
| 47641, | 1.000 AA1: | |
| 47642 | 20 AA2: 21 | MKKIFFLIILFSLLFNAAYK |
| | Probability: | |
| 72443, | 1.000 AA1: | |
| 72444 | 19 AA2: 20 | MKKIFFVICLLFSFSFCFA |
| | Probability: | |
| 15907, | 1.000 AA1: | |
| 15908 | 20 AA2: 21 | MKKIFFVVCLVFFIAGFAAA |
| | Probability: | |
| 47123, | 0.997 AA1: | |
| 47124 | 22 AA2: 23 | MKKIFIIQMMTAITLCTVCAQG |
| | Probability: | |
| 65985, | 1.000 AA1: | |
| 65986 | 25 AA2: 26 | MKKIFKIFAIIALVAVIGFSFAACS |
| | Probability: | |
| 72407, | 0.980 AA1: | |
| 72408 | 23 AA2: 24 | MKKIFKLFAIMFIFTACNIFSFG |
| | Probability: | |
| 63639, | 0.591 AA1: | |
| 63640 | 21 AA2: 22 | MKKIFKVLATQSRWLCAIALV |
| | Probability: | |
| 50335, | 0.904 AA1: | |
| 50336 | 19 AA2: 20 | MKKIFLILITVAAVFYGSG |
| | Probability: | |
| 16271, | 0.949 AA1: | |
| 16272 | 25 AA2: 26 | MKKIFRIGIAGYMGAGKSTCARSFA |
| | Probability: | |
| 3831, | 1.000 AA1: | |
| 3832 | 27 AA2: 28 | MKKIGKKHHKNLAALVIILVLSGALIA |
| | Probability: | |
| 16265, | 1.000 AA1: | |
| 16266 | 27 AA2: 28 | MKKIGKKYHKNLAALVIVLALSGALIA |
| 27233, | Probability: | MKKIGLLFLIIPVSVFLITACSRNPASG |

| | | |
|---|---|---|
| 27234 | 0.990 AA1: | |
| | 28 AA2: 29 | |
| | Probability: | |
| 56813, | 1.000 AA1: | |
| 56814 | 19 AA2: 20 | MKKIIALVIGLSIAAVCFA |
| | Probability: | |
| 61739, | 1.000 AA1: | |
| 61740 | 19 AA2: 20 | MKKIIALVIGLSISSICFA |
| | Probability: | |
| 67941, | 0.798 AA1: | |
| 67942 | 20 AA2: 21 | MKKIIERLLVFIIGIPAVYA |
| | Probability: | |
| 68753, | 0.797 AA1: | |
| 68754 | 20 AA2: 21 | MKKIIERLLVFIIGIPAVYA |
| | Probability: | |
| 54671, | 0.997 AA1: | |
| 54672 | 22 AA2: 23 | MKKIIFFLFITAAFFSSCVSMK |
| | Probability: | |
| 37329, | 0.999 AA1: | |
| 37330 | 22 AA2: 23 | MKKIIFFLFLTAAFFSSCVSMK |
| | Probability: | |
| 6061, | 0.827 AA1: | |
| 6062 | 21 AA2: 22 | MKKIIFTAIFITLVTGAFCQN |
| | Probability: | |
| 7433, | 0.646 AA1: | |
| 7434 | 22 AA2: 23 | MKKIIFTVIFITMVVTGAFCQN |
| | Probability: | |
| 51051, | 1.000 AA1: | |
| 51052 | 20 AA2: 21 | MKKIIFVVCLVFFIAGFAVA |
| | Probability: | |
| 51141, | 1.000 AA1: | |
| 51142 | 20 AA2: 21 | MKKIIIASIVILVCAGAGFA |
| 1157, | Probability: | |
| 1158 | 0.805 AA1: | MKKIIIGLGSCGIASGGLKV |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 3243, | 0.807 AA1: | |
| 3244 | 20 AA2: 21 | MKKIIIGLGSCGIASGGLKV |
| | Probability: | |
| 74819, | 1.000 AA1: | |
| 74820 | 17 AA2: 18 | MKKIIILCGLLSGFVLA |
| | Probability: | |
| 25123, | 0.936 AA1: | |
| 25124 | 26 AA2: 27 | MKKIISFCSMTVIVLCLTGCVSESHR |
| | Probability: | |
| 18621, | 0.999 AA1: | |
| 18622 | 21 AA2: 22 | MKKIITTIIVILVIIIGIAVA |
| | Probability: | |
| 22971, | 0.977 AA1: | |
| 22972 | 26 AA2: 27 | MKKIKKIINDIMFVTGAALVILTVFS |
| | Probability: | |
| 34319, | 0.990 AA1: | |
| 34320 | 25 AA2: 26 | MKKIKVFSVLIIILLICGCGSKPQS |
| | Probability: | |
| 59003, | 1.000 AA1: | |
| 59004 | 20 AA2: 21 | MKKILFCFLFAFCFPLFAQA |
| | Probability: | |
| 74791, | 1.000 AA1: | |
| 74792 | 20 AA2: 21 | MKKILLILVVLLAVSSVVFA |
| | Probability: | |
| 63831, | 0.648 AA1: | |
| 63832 | 20 AA2: 21 | MKKILLTGGGTAGHVTPNLA |
| | Probability: | |
| 56837, | 1.000 AA1: | |
| 56838 | 18 AA2: 19 | MKKILTITLLFLAGFLYA |
| | Probability: | |
| 45119, | 1.000 AA1: | |
| 45120 | 19 AA2: 20 | MKKILVFIISLSITAVCFA |

| | | |
|---|---|---|
| | Probability: | |
| 58249, | 0.913 AA1: | |
| 58250 | 26 AA2: 27 | MKKIPERGLFFIILFVLALSQLPRNA |
| | Probability: | |
| 41181, | 0.999 AA1: | |
| 41182 | 23 AA2: 24 | MKKIPVFLLLCVTLFLGCVKDNA |
| | Probability: | |
| 9409, | 0.999 AA1: | |
| 9410 | 23 AA2: 24 | MKKISIVLCVIFLLSSCATTSVK |
| | Probability: | |
| 64075, | 0.950 AA1: | |
| 64076 | 25 AA2: 26 | MKKISSRLERKKMSLCILAAICAVS |
| | Probability: | |
| 64851, | 1.000 AA1: | |
| 64852 | 19 AA2: 20 | MKKITFFLISLLISAVCFA |
| | Probability: | |
| 17515, | 1.000 AA1: | |
| 17516 | 18 AA2: 19 | MKKITLAAIACICTLSLA |
| | Probability: | |
| 58541, | 0.998 AA1: | |
| 58542 | 22 AA2: 23 | MKKIVFFLFLTAAFFSSCVSLK |
| | Probability: | |
| 65347, | 0.912 AA1: | |
| 65348 | 25 AA2: 26 | MKKIVFIALVLLICCGTFLFIGCRK |
| | Probability: | |
| 66551, | 0.857 AA1: | |
| 66552 | 18 AA2: 19 | MKKIYLLFLIFIVFSCKS |
| | Probability: | |
| 39035, | 1.000 AA1: | |
| 39036 | 20 AA2: 21 | MKKIYVLTALLIALTVSANA |
| | Probability: | |
| 66839, | 0.887 AA1: | |
| 66840 | 18 AA2: 19 | MKKKAAFVAVVGRPSAGK |
| 12481, | Probability: | MKKKEEHISYFILVFPALFLYLAVMA |

| | | |
|---|---|---|
| 12482 | 0.791 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 25649, | 0.896 AA1: | |
| 25650 | 26 AA2: 27 | MKKKEERISCFILVFPALFLYLAVMA |
| | Probability: | |
| 53523, | 0.797 AA1: | |
| 53524 | 20 AA2: 21 | MKKKILILIDLFFLTRPVVV |
| | Probability: | |
| 72409, | 0.998 AA1: | |
| 72410 | 23 AA2: 24 | MKKKINLFFILLIFSAYSLFSSG |
| | Probability: | |
| 39523, | 1.000 AA1: | |
| 39524 | 19 AA2: 20 | MKKKLIFFLLCCSLLFSSA |
| | Probability: | |
| 45327, | 0.938 AA1: | |
| 45328 | 28 AA2: 29 | MKKKLKLFLIYRTQIILLILSVSILFFS |
| | Probability: | |
| 59295, | 1.000 AA1: | |
| 59296 | 18 AA2: 19 | MKKKLLGFLLLISLPVFS |
| | Probability: | |
| 38539, | 0.618 AA1: | |
| 38540 | 22 AA2: 23 | MKKLAVVGAGVSGMTAAIYCLK |
| | Probability: | |
| 72155, | 0.993 AA1: | |
| 72156 | 19 AA2: 20 | MKKLCFFIFSAVLIFNLSA |
| | Probability: | |
| 61817, | 0.999 AA1: | |
| 61818 | 19 AA2: 20 | MKKLFGTLFALIVCLTSCA |
| | Probability: | |
| 23553, | 1.000 AA1: | |
| 23554 | 22 AA2: 23 | MKKLFKLLAIIAFIAVIGFSMA |
| 78021, | Probability: | |
| 78022 | 0.998 AA1: | MKKLFVFSMVAFSMLFLA |

| | | |
|---|---|---|
| | 18 AA2: 19 | |
| | Probability: | |
| 35859, | 0.922 AA1: | |
| 35860 | 22 AA2: 23 | MKKLHLVFIALATILFFGCSEK |
| | Probability: | |
| 27891, | 0.999 AA1: | |
| 27892 | 19 AA2: 20 | MKKLIILTLIFFSVMSVFG |
| | Probability: | |
| 19431, | 0.737 AA1: | |
| 19432 | 27 AA2: 28 | MKKLKEIFMAILVIVVVGILFYFVYQN |
| | Probability: | |
| 31813, | 0.669 AA1: | |
| 31814 | 21 AA2: 22 | MKKLKNYILVCGGTGCESSKA |
| | Probability: | |
| 44605, | 0.975 AA1: | |
| 44606 | 21 AA2: 22 | MKKLKSFFFLIPVIIIFSCAR |
| | Probability: | |
| 29551, | 0.999 AA1: | |
| 29552 | 21 AA2: 22 | MKKLLCACLVLISVLSISCQK |
| | Probability: | |
| 18399, | 0.527 AA1: | |
| 18400 | 22 AA2: 23 | MKKLLFASIFILAGLNVFCQDA |
| | Probability: | |
| 12515, | 0.996 AA1: | |
| 12516 | 19 AA2: 20 | MKKLLFVTIFILTGIYVSA |
| | Probability: | |
| 59391, | 0.713 AA1: | |
| 59392 | 13 AA2: 14 | MKKLLLICLLILA |
| | Probability: | |
| 24375, | 1.000 AA1: | |
| 24376 | 20 AA2: 21 | MKKLLLILFLIVVAGGLSFA |
| | Probability: | |
| 22921, | 1.000 AA1: | |
| 22922 | 18 AA2: 19 | MKKLLLLSSLFVALFSCA |

| | | |
|---|---|---|
| | Probability: | |
| 76585, | 1.000 AA1: | |
| 76586 | 19 AA2: 20 | MKKLLTAVLLFSVAAAAFA |
| | Probability: | |
| 8425, | 1.000 AA1: | |
| 8426 | 19 AA2: 20 | MKKLLYLFLFSMFTCPVFS |
| | Probability: | |
| 46959, | 0.896 AA1: | |
| 46960 | 19 AA2: 20 | MKKLPEKALFIIFFALAFS |
| | Probability: | |
| 18625, | 0.990 AA1: | |
| 18626 | 19 AA2: 20 | MKKLQLLGFLLLLSLPVFS |
| | Probability: | |
| 28217, | 1.000 AA1: | |
| 28218 | 21 AA2: 22 | MKKLRFIGXVLXAALCFFACA |
| | Probability: | |
| 9545, | 0.999 AA1: | |
| 9546 | 25 AA2: 26 | MKKLSKFAQVAGIMLALTFTFSCSG |
| | Probability: | |
| 33427, | 1.000 AA1: | |
| 33428 | 22 AA2: 23 | MKKLVVALAVVVLLLPGRPALS |
| | Probability: | |
| 69449, | 1.000 AA1: | |
| 69450 | 21 AA2: 22 | MKKMFFAVALLTFFVASGVNA |
| | Probability: | |
| 187, | 0.996 AA1: | |
| 188 | 25 AA2: 26 | MKKMKRKWIADVLVVVFCVSGAVFS |
| | Probability: | |
| 69659, | 1.000 AA1: | |
| 69660 | 20 AA2: 21 | MKKMPVFSIFFLILAMSASA |
| | Probability: | |
| 29519, | 1.000 AA1: | |
| 29520 | 20 AA2: 21 | MKKMPVFSIFYLVLAVSASA |
| 54829, | Probability: | MKKMVLFGSGLVVLAIVGAFLVAGCADG |

| | | |
|---|---|---|
| 54830 | 0.998 AA1: | |
| | 28 AA2: 29 | |
| | Probability: | |
| 49901, | 0.995 AA1: | |
| 49902 | 20 AA2: 21 | MKKNFSFCVIFIFVSIFIYA |
| | Probability: | |
| 62265, | 0.998 AA1: | |
| 62266 | 20 AA2: 21 | MKKNICLATIALILVVSCSS |
| | Probability: | |
| 53439, | 1.000 AA1: | |
| 53440 | 19 AA2: 20 | MKKNIIAVFWLFFAFFCYA |
| | Probability: | |
| 72199, | 0.962 AA1: | |
| 72200 | 19 AA2: 20 | MKKNILWVTAIIMMIAVIA |
| | Probability: | |
| 29857, | 1.000 AA1: | |
| 29858 | 23 AA2: 24 | MKKNLLSFVLFVAVFVIPYYAQA |
| | Probability: | |
| 46121, | 0.616 AA1: | |
| 46122 | 25 AA2: 26 | MKKNNNDLPLLFFIYFLLFIISCQQ |
| | Probability: | |
| 69553, | 0.999 AA1: | |
| 69554 | 25 AA2: 26 | MKKNYTKARKTAALYILAGLPPSNA |
| | Probability: | |
| 8461, | 0.979 AA1: | |
| 8462 | 26 AA2: 27 | MKKPFTTIIVIAVIFIGIIVAGPLFV |
| | Probability: | |
| 77161, | 1.000 AA1: | |
| 77162 | 22 AA2: 23 | MKKPIPLFFVIFLFMFSILSVA |
| | Probability: | |
| 68661, | 0.938 AA1: | |
| 68662 | 27 AA2: 28 | MKKPKKTSFAKLFTTVSLLMILTITVS |
| 52493, | Probability: | |
| 52494 | 0.997 AA1: | MKKPKKTSFAMLFITVSLLMILSITVS |

| | | |
|---|---|---|
| | 27 AA2: 28 | |
| | Probability: | |
| 23395, | 0.847 AA1: | |
| 23396 | 22 AA2: 23 | MKKPLMRSIIIVVMAAVLLELA |
| | Probability: | |
| 55947, | 0.991 AA1: | |
| 55948 | 21 AA2: 22 | MKKPLTTIIIIAVIFIGILVA |
| | Probability: | |
| 28135, | 0.967 AA1: | |
| 28136 | 29 AA2: 30 | MKKPPEKALLIVFFALVFFLFPKTALLCA |
| | Probability: | |
| 34773, | 0.999 AA1: | |
| 34774 | 29 AA2: 30 | MKKPRKTSFAMLFITVSLLMILSITVSLS |
| | Probability: | |
| 59163, | 0.997 AA1: | |
| 59164 | 23 AA2: 24 | MKKQAGICIILACCLLAACIATK |
| | Probability: | |
| 76893, | 1.000 AA1: | |
| 76894 | 24 AA2: 25 | MKKQIHSLLIASALLAIAVLGALS |
| | Probability: | |
| 33015, | 0.904 AA1: | |
| 33016 | 24 AA2: 25 | MKKQKVDLLFSKGGFVSVPPCAAA |
| | Probability: | |
| 53601, | 0.507 AA1: | |
| 53602 | 19 AA2: 20 | MKKRISILGATGSIGKSSL |
| | Probability: | |
| 51661, | 1.000 AA1: | |
| 51662 | 22 AA2: 23 | MKKRLFIGSLVFAALLAGCSLA |
| | Probability: | |
| 65183, | 0.999 AA1: | |
| 65184 | 20 AA2: 21 | MKKRLIFTFIVIASVLLFLA |
| | Probability: | |
| 45303, | 0.993 AA1: | |
| 45304 | 23 AA2: 24 | MKKRLRRAAACCLRCGNIRSSSA |

| | | |
|---|---|---|
| | Probability: | |
| 65979, | 1.000 AA1: | |
| 65980 | 26 AA2: 27 | MKKRLTKLAQVAGIMLALAFTFSCSS |
| | Probability: | |
| 57781, | 1.000 AA1: | |
| 57782 | 29 AA2: 30 | MKKRMYLLCGIFAAVMIIAAVFTSCVTPA |
| | Probability: | |
| 11905, | 0.525 AA1: | |
| 11906 | 19 AA2: 20 | MKKRSVXAGPALIDAVALM |
| | Probability: | |
| 11035, | 0.997 AA1: | |
| 11036 | 22 AA2: 23 | MKKRTKPALAALILSVIMILAG |
| | Probability: | |
| 19449, | 0.962 AA1: | |
| 19450 | 22 AA2: 23 | MKKRVLLIAFICGLMFTTCDLL |
| | Probability: | |
| 73435, | 0.998 AA1: | |
| 73436 | 20 AA2: 21 | MKKRVYLIAAVVSGALAVSG |
| | Probability: | |
| 31345, | 1.000 AA1: | |
| 31346 | 19 AA2: 20 | MKKSAFFLLFLVLAFALLA |
| | Probability: | |
| 8865, | 0.999 AA1: | |
| 8866 | 21 AA2: 22 | MKKSIFITLFVITFAFQPLFA |
| | Probability: | |
| 42111, | 0.999 AA1: | |
| 42112 | 21 AA2: 22 | MKKSILAVMVLYLTLHTTLNA |
| | Probability: | |
| 6451, | 0.991 AA1: | |
| 6452 | 20 AA2: 21 | MKKSKYLSFFLFSAGLLILA |
| | Probability: | |
| 72573, | 0.986 AA1: | |
| 72574 | 20 AA2: 21 | MKKSKYLSFFLFSVGLLIFA |
| 17955, | Probability: | MKKSLFIFIALVLTAAMAFA |

| | | |
|---|---|---|
| 17956 | 1.000 AA1: | |
| | 20 AA2: 21 | |
| | Probability: | |
| 64373, | 1.000 AA1: | |
| 64374 | 25 AA2: 26 | MKKSLFIFMSFVLTAALAFAGGGQA |
| | Probability: | |
| 57309, | 1.000 AA1: | |
| 57310 | 20 AA2: 21 | MKKTAFLALIFYAIAANGYA |
| | Probability: | |
| 46469, | 0.988 AA1: | |
| 46470 | 21 AA2: 22 | MKKTAPLILVIFILLTGCATT |
| | Probability: | |
| 40881, | 0.998 AA1: | |
| 40882 | 20 AA2: 21 | MKKTFFFLFLVFSFYGVCFS |
| | Probability: | |
| 11953, | 0.992 AA1: | |
| 11954 | 20 AA2: 21 | MKKTGKFLGLLVIVAVIGFA |
| | Probability: | |
| 55107, | 0.972 AA1: | |
| 55108 | 19 AA2: 20 | MKKTIILSSVVILIGLLIA |
| | Probability: | |
| 51053, | 0.979 AA1: | |
| 51054 | 28 AA2: 29 | MKKTIMLALMVMIFFITIITCKHVDSLA |
| | Probability: | |
| 42875, | 1.000 AA1: | |
| 42876 | 24 AA2: 25 | MKKTIVSLLIAGALVLAFNHTVEA |
| | Probability: | |
| 55105, | 1.000 AA1: | |
| 55106 | 18 AA2: 19 | MKKTLFLLFLFLSAVVFA |
| | Probability: | |
| 63627, | 0.861 AA1: | |
| 63628 | 22 AA2: 23 | MKKTPAPIGLIISLCIIGMFMN |
| 66113, | Probability: | |
| 66114 | 0.574 AA1: | MKKTPILITALVFIVSCYQ |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 27503, | 1.000 AA1: | |
| 27504 | 22 AA2: 23 | MKKTTAILFVLLAVCFTAQVYA |
| | Probability: | |
| 4321, | 1.000 AA1: | |
| 4322 | 19 AA2: 20 | MKKTVFLVAACFLTTSLYS |
| | Probability: | |
| 64327, | 0.999 AA1: | |
| 64328 | 21 AA2: 22 | MKKTVLLILTIFILLPSCATS |
| | Probability: | |
| 5433, | 0.920 AA1: | |
| 5434 | 21 AA2: 22 | MKKTVLLVVLFAVITALIYIG |
| | Probability: | |
| 56953, | 0.957 AA1: | |
| 56954 | 21 AA2: 22 | MKKTVTTIVVIFVIFLGILVA |
| | Probability: | |
| 18405, | 0.965 AA1: | |
| 18406 | 25 AA2: 26 | MKKVALIAALIFIVTIVMLALYCRT |
| | Probability: | |
| 53821, | 0.991 AA1: | |
| 53822 | 22 AA2: 23 | MKKVFLLLIMVIPAALIVSCSG |
| | Probability: | |
| 5139, | 0.974 AA1: | |
| 5140 | 24 AA2: 25 | MKKVGLVLAILMVAIILFSTGCKK |
| | Probability: | |
| 67703, | 1.000 AA1: | |
| 67704 | 22 AA2: 23 | MKKVIIFFILSFSFCGFCFSQA |
| | Probability: | |
| 4665, | 1.000 AA1: | |
| 4666 | 22 AA2: 23 | MKKVIIFLVLVFAFCGLCFSQA |
| | Probability: | |
| 26881, | 0.802 AA1: | |
| 26882 | 19 AA2: 20 | MKKVIISSAIVIVLGLLIA |

| | | |
|---|---|---|
| | Probability: | |
| 40235, | 1.000 AA1: | |
| 40236 | 20 AA2: 21 | MKKVITSLLISAALATVAFA |
| | Probability: | |
| 40073, | 0.932 AA1: | |
| 40074 | 20 AA2: 21 | MKKVLFLLFLSFIAGDEAWA |
| | Probability: | |
| 24695, | 0.826 AA1: | |
| 24696 | 25 AA2: 26 | MKKVLRIGIAGYMGAGKSTCARSLA |
| | Probability: | |
| 38639, | 0.950 AA1: | |
| 38640 | 25 AA2: 26 | MKKVLRIGLAGYMGAGKSTCARSFA |
| | Probability: | |
| 74365, | 1.000 AA1: | |
| 74366 | 23 AA2: 24 | MKKVLSGFAFFSLFALLMLSCAG |
| | Probability: | |
| 12309, | 0.947 AA1: | |
| 12310 | 23 AA2: 24 | MKKVVFCLLSITSFTIMLFLLHG |
| | Probability: | |
| 77863, | 0.998 AA1: | |
| 77864 | 22 AA2: 23 | MKKWLKISLISLASLIGVFLLA |
| | Probability: | |
| 4653, | 1.000 AA1: | |
| 4654 | 19 AA2: 20 | MKKYALLALLLYTAVSAFA |
| | Probability: | |
| 34709, | 1.000 AA1: | |
| 34710 | 23 AA2: 24 | MKKYILALFILLFVVVFSAAQTA |
| | Probability: | |
| 72139, | 0.667 AA1: | |
| 72140 | 26 AA2: 27 | MKKYINHLLLPIIATVLMFIGCGKSN |
| | Probability: | |
| 6319, | 1.000 AA1: | |
| 6320 | 20 AA2: 21 | MKKYLIVFCLVFFIAGFAAA |
| 47301, | Probability: | MKKYLKQAVILFILLFICISCSG |

| | | |
|---|---|---|
| 47302 | 0.977 AA1: | |
| | 23 AA2: 24 | |
| | Probability: | |
| 53633, | 0.999 AA1: | |
| 53634 | 20 AA2: 21 | MKKYPVFLLFSFIITFTLNA |
| | Probability: | |
| 59307, | 0.974 AA1: | |
| 59308 | 26 AA2: 27 | MKKYSVLFVVFFFTLGLWLQPLPLFS |
| | Probability: | |
| 32679, | 1.000 AA1: | |
| 32680 | 21 AA2: 22 | MKKYTLFFICCLMSLMXVVTA |
| | Probability: | |
| 35107, | 0.999 AA1: | |
| 35108 | 21 AA2: 22 | MKKYTLFFICCLVSLIVVVTA |
| | Probability: | |
| 47829, | 1.000 AA1: | |
| 47830 | 23 AA2: 24 | MKKYTLFFVFCLISLVAVAAVSA |
| | Probability: | |
| 62509, | 0.999 AA1: | |
| 62510 | 23 AA2: 24 | MKKYTLLFICLLIPFIMAITACS |
| | Probability: | |
| 25987, | 0.953 AA1: | |
| 25988 | 23 AA2: 24 | MKLAFFKIVLLTVFSLMFIQCHG |
| | Probability: | |
| 39105, | 0.709 AA1: | |
| 39106 | 27 AA2: 28 | MKLALIIGIGSFIGGFSRYLLSLLIQA |
| | Probability: | |
| 25557, | 0.933 AA1: | |
| 25558 | 19 AA2: 20 | MKLFEIILIAVGLSMDAFA |
| | Probability: | |
| 62915, | 0.520 AA1: | |
| 62916 | 26 AA2: 27 | MKLFTLPRLVPRNNYLQLLFVFAAFA |
| 31987, | Probability: | |
| 31988 | 0.689 AA1: | MKLFTLPRLVPRNNYLQLLFVFAAFALM |

| | | |
|---|---|---|
| | 28 AA2: 29 | |
| | Probability: | |
| 19907, | 0.907 AA1: | |
| 19908 | 29 AA2: 30 | MKLIELTKNCFFLAFLTLLLSIGTFYTFK |
| | Probability: | |
| 9581, | 0.658 AA1: | |
| 9582 | 18 AA2: 19 | MKLIKKFLGIAAIIAVTG |
| | Probability: | |
| 35007, | 0.998 AA1: | |
| 35008 | 24 AA2: 25 | MKLKHKLPLILFIAFVAVISLTFT |
| | Probability: | |
| 63299, | 0.883 AA1: | |
| 63300 | 25 AA2: 26 | MKLKKPLLKIVASPLLGPVIAIGSA |
| | Probability: | |
| 73711, | 0.518 AA1: | |
| 73712 | 26 AA2: 27 | MKLKLYLRQTNANPSVANAGITACLA |
| | Probability: | |
| 59685, | 1.000 AA1: | |
| 59686 | 22 AA2: 23 | MKLKNIACLSLFLLFVSTPLLT |
| | Probability: | |
| 30845, | 0.744 AA1: | |
| 30846 | 23 AA2: 24 | MKLKTLLIGMAPAVLVSILCVFG |
| | Probability: | |
| 70431, | 0.999 AA1: | |
| 70432 | 26 AA2: 27 | MKLKTMKLAAIIAALFLAIGCAGSPK |
| | Probability: | |
| 39213, | 0.997 AA1: | |
| 39214 | 20 AA2: 21 | MKLKYIAFALLVALAVSCIS |
| | Probability: | |
| 21523, | 0.940 AA1: | |
| 21524 | 27 AA2: 28 | MKLLKENGARKVICSISLPLFSGSAVS |
| | Probability: | |
| 15737, | 0.997 AA1: | |
| 15738 | 18 AA2: 19 | MKLLKFFSAILVSLFVVA |

| | | |
|---|---|---|
| | Probability: | |
| 30873, | 0.983 AA1: | |
| 30874 | 29 AA2: 30 | MKLLKNLLKSPMFVIGASLFLVTLLIALV |
| | Probability: | |
| 72053, | 0.951 AA1: | |
| 72054 | 26 AA2: 27 | MKLLKYVIGLLLAVLGLYVFFHGSGG |
| | Probability: | |
| 8421, | 0.845 AA1: | |
| 8422 | 21 AA2: 22 | MKLLLHACCAPCSIQCVESLA |
| | Probability: | |
| 34035, | 0.986 AA1: | |
| 34036 | 21 AA2: 22 | MKLLLHCCCAPCSSSCLPALR |
| | Probability: | |
| 69849, | 0.944 AA1: | |
| 69850 | 22 AA2: 23 | MKLLLHCCCAPCSVSCIRSLRA |
| | Probability: | |
| 38569, | 0.683 AA1: | |
| 38570 | 21 AA2: 22 | MKLLLHTCCAPCSIQCVESLA |
| | Probability: | |
| 20931, | 0.996 AA1: | |
| 20932 | 17 AA2: 18 | MKLLVVLILIFFSLANS |
| | Probability: | |
| 11789, | 1.000 AA1: | |
| 11790 | 22 AA2: 23 | MKLNAKLLVFGACFLMARPLFA |
| | Probability: | |
| 12491, | 0.804 AA1: | |
| 12492 | 23 AA2: 24 | MKLNGELAFAVVIAILIVTLALL |
| | Probability: | |
| 42025, | 1.000 AA1: | |
| 42026 | 22 AA2: 23 | MKLNGKLLVFGAFLLMARPLFA |
| | Probability: | |
| 1509, | 0.959 AA1: | |
| 1510 | 26 AA2: 27 | MKLNKIFILFAAAAIILGILNSYFIM |
| 15479, | Probability: | MKLNSPFLNNLGFIALVFFACSAFS |

| | | |
|---|---|---|
| 15480 | 1.000 AA1: | |
| | 25 AA2: 26 | |
| | Probability: | |
| 58751, | 1.000 AA1: | |
| 58752 | 25 AA2: 26 | MKLNSPFLNNLGFIALVFFACSAFS |
| | Probability: | |
| 46505, | 1.000 AA1: | |
| 46506 | 22 AA2: 23 | MKLNTRKLIALFLAVFAITSHA |
| | Probability: | |
| 16273, | 1.000 AA1: | |
| 16274 | 22 AA2: 23 | MKLNVKFLALGACLLMARPLFA |
| | Probability: | |
| 71381, | 1.000 AA1: | |
| 71382 | 22 AA2: 23 | MKLQTLVITAVLAVFIQQAAFA |
| | Probability: | |
| 52921, | 0.509 AA1: | |
| 52922 | 23 AA2: 24 | MKLRGNTAMGLALMAIQKGYSLK |
| | Probability: | |
| 15611, | 1.000 AA1: | |
| 15612 | 25 AA2: 26 | MKLRLLSXSSLALLCLLITVCPTQA |
| | Probability: | |
| 69755, | 0.999 AA1: | |
| 69756 | 25 AA2: 26 | MKLSAGLFLRTAAVAVCVAVGAAFG |
| | Probability: | |
| 68159, | 0.987 AA1: | |
| 68160 | 25 AA2: 26 | MKLSAYISAALLACSPLAFFRPVQV |
| | Probability: | |
| 49043, | 0.979 AA1: | |
| 49044 | 26 AA2: 27 | MKLSENNGFKLSLFILILAGLSPLCA |
| | Probability: | |
| 36617, | 1.000 AA1: | |
| 36618 | 27 AA2: 28 | MKLSGGDMRLFRLFAAFLLLSAVLAWA |
| 71309, | Probability: | |
| 71310 | 0.972 AA1: | MKLSIRVPLLIVAVVLITSA |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 56737, | 0.995 AA1: | |
| 56738 | 27 AA2: 28 | MKLSQMSIKSKMLVFSGLFFMTVAVLG |
| | Probability: | |
| 64375, | 0.989 AA1: | |
| 64376 | 22 AA2: 23 | MKLTALAFLGVIGLFIVTPLFS |
| | Probability: | |
| 4791, | 0.579 AA1: | |
| 4792 | 25 AA2: 26 | MKLTGAVMKGPILAVLAAGMGSRYG |
| | Probability: | |
| 54207, | 0.949 AA1: | |
| 54208 | 17 AA2: 18 | MKLTSILFCLFALSCSS |
| | Probability: | |
| 38805, | 0.964 AA1: | |
| 38806 | 24 AA2: 25 | MKLYSRGQVLFFSFLSAVIVLLFA |
| | Probability: | |
| 25839, | 0.980 AA1: | |
| 25840 | 26 AA2: 27 | MKMGTRKRAKMKAKPAAVAAIVIAVA |
| | Probability: | |
| 66909, | 0.722 AA1: | |
| 66910 | 18 AA2: 19 | MKMKLGSLFDGIGGFSLA |
| | Probability: | |
| 70433, | 0.994 AA1: | |
| 70434 | 25 AA2: 26 | MKMKLTAAIIAAFVIIMGCAGGPKA |
| | Probability: | |
| 56981, | 0.999 AA1: | |
| 56982 | 27 AA2: 28 | MKMKLTAAIIAAFVIIMGCAGGPKASK |
| | Probability: | |
| 45971, | 1.000 AA1: | |
| 45972 | 25 AA2: 26 | MKMKNKFFMFLLTASFLFLSFPLFA |
| | Probability: | |
| 41589, | 1.000 AA1: | |
| 41590 | 26 AA2: 27 | MKMKNKYTILFTLFILSILVQTCASV |

| | | |
|---|---|---|
| | Probability: | |
| 16863, | 0.996 AA1: | |
| 16864 | 25 AA2: 26 | MKMKPDFSLFLLQVVILIIVSAAFA |
| | Probability: | |
| 52987, | 0.998 AA1: | |
| 52988 | 25 AA2: 26 | MKMNRKLFIALATLLLFGMLAGCRK |
| | Probability: | |
| 77659, | 1.000 AA1: | |
| 77660 | 20 AA2: 21 | MKMNVFVAVLFVGFVFSVFS |
| | Probability: | |
| 43971, | 0.974 AA1: | |
| 43972 | 26 AA2: 27 | MKMPKNYFILILKVLMLGVFMLIAKA |
| | Probability: | |
| 32917, | 0.998 AA1: | |
| 32918 | 30 AA2: 31 | MKMRFKLSLWLIAVMAVVVTIVATVLIRYA |
| | Probability: | |
| 76349, | 0.826 AA1: | |
| 76350 | 15 AA2: 16 | MKMSAQLAAFHPAVA |
| | Probability: | |
| 10641, | 0.917 AA1: | |
| 10642 | 21 AA2: 22 | MKMTRKLCLCPPGLLMESQRA |
| | Probability: | |
| 37147, | 0.768 AA1: | |
| 37148 | 24 AA2: 25 | MKNAHAEHIAQTVATVLSLGGANA |
| | Probability: | |
| 44559, | 0.923 AA1: | |
| 44560 | 20 AA2: 21 | MKNCIFICIFISFINLGLYS |
| | Probability: | |
| 46045, | 0.913 AA1: | |
| 46046 | 22 AA2: 23 | MKNEKGKNKFRILSFTLLIILA |
| | Probability: | |
| 37029, | 1.000 AA1: | |
| 37030 | 22 AA2: 23 | MKNFCKLITVLILFAAFSPLFA |
| 55305, | Probability: | MKNFIFAVLIVFLFLSVVFPVQA |

| | | |
|---|---|---|
| 55306 | 1.000 AA1: | |
| | 23 AA2: 24 | |
| | Probability: | |
| 28689, | 0.745 AA1: | |
| 28690 | 24 AA2: 25 | MKNFMLCLLNVGLLVCGQLLFKLA |
| | Probability: | |
| 56477, | 0.897 AA1: | |
| 56478 | 21 AA2: 22 | MKNFRRVLLFVLLFGTMLIFG |
| | Probability: | |
| 43785, | 0.956 AA1: | |
| 43786 | 18 AA2: 19 | MKNFVFICAFVSVFSSCS |
| | Probability: | |
| 38611, | 0.947 AA1: | |
| 38612 | 28 AA2: 29 | MKNIAGMVKLGCILAVYAAVACVMLAFV |
| | Probability: | |
| 47375, | 1.000 AA1: | |
| 47376 | 20 AA2: 21 | MKNIAVLSVLVIILSLNSYA |
| | Probability: | |
| 70709, | 1.000 AA1: | |
| 70710 | 25 AA2: 26 | MKNIFKLFGFIALAAIIVFSMAACG |
| | Probability: | |
| 71373, | 1.000 AA1: | |
| 71374 | 19 AA2: 20 | MKNIGFILLLLLTAQMVIA |
| | Probability: | |
| 6553, | 1.000 AA1: | |
| 6554 | 19 AA2: 20 | MKNIGLFLLLLIIAQMVIA |
| | Probability: | |
| 53669, | 0.978 AA1: | |
| 53670 | 26 AA2: 27 | MKNIIKASIIFSAIFFFVKCSFSSPA |
| | Probability: | |
| 46787, | 1.000 AA1: | |
| 46788 | 25 AA2: 26 | MKNIIKFRITGFLTMLAVIAAFAAA |
| 52859, | Probability: | |
| 52860 | 0.999 AA1: | MKNIIKQICVIAFVAIIGFSFFA |

| | | |
|---|---|---|
| | 23 AA2: 24 | |
| | Probability: | |
| 21571, | 1.000 AA1: | |
| 21572 | 26 AA2: 27 | MKNIIKSRITGFLTMLAVIAAFGAAA |
| | Probability: | |
| 74739, | 0.999 AA1: | |
| 74740 | 26 AA2: 27 | MKNIIRKNAYKIFLFLLAVIPASISG |
| | Probability: | |
| 40603, | 0.994 AA1: | |
| 40604 | 27 AA2: 28 | MKNIKDKLILITGGASGIGQLMAFAFA |
| | Probability: | |
| 24737, | 0.993 AA1: | |
| 24738 | 25 AA2: 26 | MKNIKKVFAVIAVTIIVVTGMAVIG |
| | Probability: | |
| 18517, | 0.996 AA1: | |
| 18518 | 25 AA2: 26 | MKNIMKVFKFIAITIIIVLVMAVLG |
| | Probability: | |
| 36509, | 0.757 AA1: | |
| 36510 | 18 AA2: 19 | MKNKFFICFLFSIFSFIS |
| | Probability: | |
| 37609, | 1.000 AA1: | |
| 37610 | 26 AA2: 27 | MKNKFSKLVLSASIMLAMVFTFSCSG |
| | Probability: | |
| | 0.888 AA1: | |
| 93, 94 | 30 AA2: 31 | MKNKINVLSVRVFGIIAIVAVIGFSMIACN |
| | Probability: | |
| 8727, | 1.000 AA1: | |
| 8728 | 20 AA2: 21 | MKNKKIIFLLLIFFIFSCQS |
| | Probability: | |
| 18627, | 0.990 AA1: | |
| 18628 | 22 AA2: 23 | MKNKLFLAGIIIILIFGIMVTG |
| | Probability: | |
| 41347, | 1.000 AA1: | |
| 41348 | 24 AA2: 25 | MKNKLFTILMLLFALSFASCASMA |

| | | |
|---|---|---|
| | Probability: | |
| 70915, | 1.000 AA1: | |
| 70916 | 23 AA2: 24 | MKNKLIACIGLVFMFLSVISVFA |
| | Probability: | |
| 61577, | 0.999 AA1: | |
| 61578 | 20 AA2: 21 | MKNKLLFFVILISFVFSCYS |
| | Probability: | |
| 18155, | 0.775 AA1: | |
| 18156 | 22 AA2: 23 | MKNKLSGVIHPAVIAVWAAVCA |
| | Probability: | |
| 61379, | 1.000 AA1: | |
| 61380 | 23 AA2: 24 | MKNKLWVVAILTAFTLSTTYCFA |
| | Probability: | |
| 9651, | 0.933 AA1: | |
| 9652 | 26 AA2: 27 | MKNKSPRIFLKIFIYTVFLVLLICLT |
| | Probability: | |
| 6965, | 0.970 AA1: | |
| 6966 | 20 AA2: 21 | MKNKYIFFLAVLLVFFNLWS |
| | Probability: | |
| 41287, | 1.000 AA1: | |
| 41288 | 23 AA2: 24 | MKNKYIILPLIFFCFSASFTVFA |
| | Probability: | |
| 56749, | 0.989 AA1: | |
| 56750 | 23 AA2: 24 | MKNLCVFLGVLLLSACQSINRNN |
| | Probability: | |
| 251, | 0.999 AA1: | |
| 252 | 23 AA2: 24 | MKNLFKTAFILIFAAAGSLPVNS |
| | Probability: | |
| 21649, | 0.806 AA1: | |
| 21650 | 27 AA2: 28 | MKNLIITILIIFSGTIFSCGHISSFFN |
| | Probability: | |
| 34413, | 0.957 AA1: | |
| 34414 | 28 AA2: 29 | MKNLKVSAKLLVGFGVSAAFLLMVGVVS |
| 37161, | Probability: | MKNLKVSMKLLVGFGLMVALILTISLV |

| | | |
|---|---|---|
| 37162 | 0.969 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 52387, | 1.000 AA1: | |
| 52388 | 19 AA2: 20 | MKNLLLALFTLAFAGSVMA |
| | Probability: | |
| 32477, | 0.995 AA1: | |
| 32478 | 26 AA2: 27 | MKNLLRLTMTALFCLLIPPAFWPAQA |
| | Probability: | |
| 41323, | 0.921 AA1: | |
| 41324 | 22 AA2: 23 | MKNLLRNLKATGPAFLIGVALS |
| | Probability: | |
| 13903, | 0.945 AA1: | |
| 13904 | 26 AA2: 27 | MKNLSASIXSLVLAIIFCYMIIASIN |
| | Probability: | |
| 7925, | 0.918 AA1: | |
| 7926 | 25 AA2: 26 | MKNLVKTLVILNAVIFAACGNNSST |
| | Probability: | |
| 71073, | 0.996 AA1: | |
| 71074 | 25 AA2: 26 | MKNNKKLILALSAILIVTAIFIVSE |
| | Probability: | |
| 49721, | 0.999 AA1: | |
| 49722 | 25 AA2: 26 | MKNNKKLVLALSAILIIAAIFIVSE |
| | Probability: | |
| 69379, | 1.000 AA1: | |
| 69380 | 26 AA2: 27 | MKNNLFFILLLIFTISFVISCSSSDA |
| | Probability: | |
| 661, | 1.000 AA1: | |
| 662 | 28 AA2: 29 | MKNNLFSILLLIFTISFVISCSSSSVNG |
| | Probability: | |
| 12757, | 0.847 AA1: | |
| 12758 | 22 AA2: 23 | MKNNRNSLVATMGFALIFPGGG |
| 38661, | Probability: | |
| 38662 | 0.978 AA1: | MKNPVFLCAIVSVXSSXS |

| | | |
|---|---|---|
| | 18 AA2: 19 | |
| | Probability: | |
| 9341, | 1.000 AA1: | |
| 9342 | 25 AA2: 26 | MKNQLKLLRLAIMVMAIFFSLPFTA |
| | Probability: | |
| 34355, | 0.953 AA1: | |
| 34356 | 25 AA2: 26 | MKNRIRFLWFIVLATVIVISFPSCK |
| | Probability: | |
| 4561, | 1.000 AA1: | |
| 4562 | 21 AA2: 22 | MKNRIVVLLYLLALISRPLFA |
| | Probability: | |
| 44115, | 0.726 AA1: | |
| 44116 | 18 AA2: 19 | MKNRKTKIVCSLGPASSG |
| | Probability: | |
| 23495, | 0.886 AA1: | |
| 23496 | 19 AA2: 20 | MKNSPIILGITGASGAIYA |
| | Probability: | |
| 29897, | 0.999 AA1: | |
| 29898 | 21 AA2: 22 | MKNTAKLLLASLALGAAVSCG |
| | Probability: | |
| 55143, | 0.680 AA1: | |
| 55144 | 23 AA2: 24 | MKNTIGNRRGFLPILFGMFVITA |
| | Probability: | |
| 25635, | 0.993 AA1: | |
| 25636 | 23 AA2: 24 | MKNTIKFRFAGIGLLLVTVAVFS |
| | Probability: | |
| 64371, | 1.000 AA1: | |
| 64372 | 25 AA2: 26 | MKNTIKFRFMGILTLLVTCAIFAVA |
| | Probability: | |
| 10565, | 1.000 AA1: | |
| 10566 | 25 AA2: 26 | MKNTIKFRFMGILTMLVICAVFAVA |
| | Probability: | |
| 73859, | 0.996 AA1: | |
| 73860 | 23 AA2: 24 | MKNTITIAIMILFFALFQSTVFV |

| | | |
|---|---|---|
| | Probability: | |
| 34539, | 1.000 AA1: | |
| 34540 | 24 AA2: 25 | MKNTRKTKIILLLAAALGLTSVFA |
| | Probability: | |
| 75345, | 1.000 AA1: | |
| 75346 | 24 AA2: 25 | MKNVFLTSFLLIXCAALYAPALYA |
| | Probability: | |
| 48349, | 0.967 AA1: | |
| 48350 | 24 AA2: 25 | MKNVKIGAKLIMSFLFLAALTAFM |
| | Probability: | |
| 2457, | 0.582 AA1: | |
| 2458 | 30 AA2: 31 | MKNVPILFWKRLRGLFLVVLFVLFFFQLPK |
| | Probability: | |
| 63937, | 0.953 AA1: | |
| 63938 | 16 AA2: 17 | MKNWSWACASATIVSA |
| | Probability: | |
| 33269, | 0.978 AA1: | |
| 33270 | 21 AA2: 22 | MKNYPVLFVLFFFSLGFPLFS |
| | Probability: | |
| 53789, | 0.968 AA1: | |
| 53790 | 16 AA2: 17 | MKPFFALIFMVTLVNA |
| | Probability: | |
| 63833, | 0.999 AA1: | |
| 63834 | 26 AA2: 27 | MKPFPLVKIFLFFLGLFFSVPAVLQA |
| | Probability: | |
| 22531, | 0.998 AA1: | |
| 22532 | 27 AA2: 28 | MKPIIHKTPLFLALLTALLGPLGPVQA |
| | Probability: | |
| 60311, | 0.986 AA1: | |
| 60312 | 28 AA2: 29 | MKPKFVIRRITILLLLTIVLALLSGCNR |
| | Probability: | |
| 48473, | 0.998 AA1: | |
| 48474 | 27 AA2: 28 | MKPPFIPRFAALLLYAVSMLISPAALR |
| 931, | Probability: | MKQFKEIVLGIFVILVFGYAIT |

| | | |
|---|---|---|
| 932 | 0.554 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 42855, | 1.000 AA1: | |
| 42856 | 23 AA2: 24 | MKQKKLISFLTALLFLILFTACA |
| | Probability: | |
| 73773, | 0.745 AA1: | |
| 73774 | 28 AA2: 29 | MKQSAFKLSYKTILNFLLIIVLLGGSVS |
| | Probability: | |
| 22499, | 0.992 AA1: | |
| 22500 | 21 AA2: 22 | MKQYLFLTILILTGMWVSAQN |
| | Probability: | |
| 111, | 0.996 AA1: | |
| 112 | 24 AA2: 25 | MKRAFFLTALIICIMLAVSCKSGS |
| | Probability: | |
| 44619, | 0.979 AA1: | |
| 44620 | 24 AA2: 25 | MKRCTAIFVTVATAVLFLSSCKQS |
| | Probability: | |
| 62361, | 0.983 AA1: | |
| 62362 | 26 AA2: 27 | MKREEARAIAQAPAAAAPTAPPPVAA |
| | Probability: | |
| 54653, | 1.000 AA1: | |
| 54654 | 21 AA2: 22 | MKRFAAGSLLAILLFAVSAIA |
| | Probability: | |
| 27501, | 1.000 AA1: | |
| 27502 | 24 AA2: 25 | MKRFAFLSMLVLLGLLCACTAANK |
| | Probability: | |
| 37133, | 1.000 AA1: | |
| 37134 | 20 AA2: 21 | MKRFLPALILWWCCCAPCFA |
| | Probability: | |
| 69269, | 0.537 AA1: | |
| 69270 | 20 AA2: 21 | MKRFYVYLLIIFLGMTVFYA |
| 40541, | Probability: | |
| 40542 | 1.000 AA1: | MKRGFFCITALFFFVSVSAFA |

| | | |
|---|---|---|
| | 21 AA2: 22 | |
| | Probability: | |
| 61601, | 1.000 AA1: | |
| 61602 | 22 AA2: 23 | MKRIAFFTFFSLFLTSLCFAQA |
| | Probability: | |
| 29485, | 0.997 AA1: | |
| 29486 | 19 AA2: 20 | MKRIFRWIFAGIVAIPVMA |
| | Probability: | |
| 43941, | 0.998 AA1: | |
| 43942 | 20 AA2: 21 | MKRIFTFILFTLIAVFPLAA |
| | Probability: | |
| 71841, | 0.997 AA1: | |
| 71842 | 20 AA2: 21 | MKRIFTFVIFTLIAVFPLSA |
| | Probability: | |
| 12781, | 1.000 AA1: | |
| 12782 | 20 AA2: 21 | MKRIGFFLIVVLMSAPFVNA |
| | Probability: | |
| 5259, | 0.987 AA1: | |
| 5260 | 21 AA2: 22 | MKRIGFIVLALLLCGAMVLTG |
| | Probability: | |
| 45955, | 0.971 AA1: | |
| 45956 | 21 AA2: 22 | MKRIIFAAFFITLAAGAFCQN |
| | Probability: | |
| 33757, | 1.000 AA1: | |
| 33758 | 19 AA2: 20 | MKRIIFFLLGLFIALPCFS |
| | Probability: | |
| 68605, | 1.000 AA1: | |
| 68606 | 19 AA2: 20 | MKRIIFFLLGLFIALPCFS |
| | Probability: | |
| 71847, | 0.991 AA1: | |
| 71848 | 19 AA2: 20 | MKRIIFLLTLTAAVLFTLG |
| | Probability: | |
| 60107, | 0.991 AA1: | |
| 60108 | 25 AA2: 26 | MKRIIRFSKYFIPAALLSSTIAVFS |

| | | |
|---|---|---|
| | Probability: | |
| 57859, | 1.000 AA1: | |
| 57860 | 22 AA2: 23 | MKRIKFFLLLTLIFSFMPVTVG |
| | Probability: | |
| 69769, | 0.996 AA1: | |
| 69770 | 23 AA2: 24 | MKRKIIWLILFALVLLSSCKFFL |
| | Probability: | |
| 49241, | 0.694 AA1: | |
| 49242 | 15 AA2: 16 | MKRKLAAINASLSLA |
| | Probability: | |
| 64341, | 1.000 AA1: | |
| 64342 | 19 AA2: 20 | MKRKLFFILTTVVLLIATA |
| | Probability: | |
| 67249, | 0.994 AA1: | |
| 67250 | 26 AA2: 27 | MKRKRKRKMKFSKQVLTALTAIVIVA |
| | Probability: | |
| 71619, | 1.000 AA1: | |
| 71620 | 21 AA2: 22 | MKRKVFLVLALFALMAAAVFA |
| | Probability: | |
| 40907, | 0.885 AA1: | |
| 40908 | 30 AA2: 31 | MKRKVKKAMTYPAMVAVVAILVVIIMLTFV |
| | Probability: | |
| 57379, | 1.000 AA1: | |
| 57380 | 21 AA2: 22 | MKRKVLFVFTLLLVLGMAAYA |
| | Probability: | |
| 63127, | 0.999 AA1: | |
| 63128 | 27 AA2: 28 | MKRLATARFLAIIIGGMTSLMIPSLVA |
| | Probability: | |
| 10743, | 0.984 AA1: | |
| 10744 | 20 AA2: 21 | MKRLFAFFVVIVICSCAAFT |
| | Probability: | |
| 59827, | 0.999 AA1: | |
| 59828 | 17 AA2: 18 | MKRLFLLLIIIPSFVFA |
| 55521, | Probability: | MKRLFLLLLLIPSFVFA |

| | | |
|---|---|---|
| 55522 | 1.000 AA1: | |
| | 17 AA2: 18 | |
| | Probability: | |
| 43625, | 0.624 AA1: | |
| 43626 | 20 AA2: 21 | MKRLFVFPAWILIFSLTILS |
| | Probability: | |
| 3167, | 0.998 AA1: | |
| 3168 | 20 AA2: 21 | MKRLLLVIFAGLLFNVHVFA |
| | Probability: | |
| 66835, | 0.998 AA1: | |
| 66836 | 19 AA2: 20 | MKRLLSVILIALLCIAALA |
| | Probability: | |
| 61763, | 0.977 AA1: | |
| 61764 | 26 AA2: 27 | MKRLPTVRFLAIIIGGMTSLMIPSFV |
| | Probability: | |
| 74703, | 0.985 AA1: | |
| 74704 | 23 AA2: 24 | MKRLQDKVAIITGGAAGIGAATA |
| | Probability: | |
| 46285, | 1.000 AA1: | |
| 46286 | 20 AA2: 21 | MKRLVFVLALVFIASAFIFA |
| | Probability: | |
| 54009, | 1.000 AA1: | |
| 54010 | 21 AA2: 22 | MKRMSFAVSLFTFLFAVSAYS |
| | Probability: | |
| 2573, | 0.998 AA1: | |
| 2574 | 24 AA2: 25 | MKRNFCTAVVSLAMILGGLQTVCA |
| | Probability: | |
| 8045, | 1.000 AA1: | |
| 8046 | 21 AA2: 22 | MKRNIFLFLFCFLIFSAVLYG |
| | Probability: | |
| 24963, | 1.000 AA1: | |
| 24964 | 24 AA2: 25 | MKRNNSASFLIAVSFSILFTACVA |
| 38667, | Probability: | |
| 38668 | 0.993 AA1: | MKRNTFIFILMTVVAILALGSCVLT |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 72165, | 0.998 AA1: | |
| 72166 | 23 AA2: 24 | MKRNVIFGILFFVMIMTTFGSQT |
| | Probability: | |
| 74579, | 1.000 AA1: | |
| 74580 | 22 AA2: 23 | MKRQPSLASAVALLTVAFFLVG |
| | Probability: | |
| 40837, | 0.965 AA1: | |
| 40838 | 20 AA2: 21 | MKRRFAVFFFFTFVIPCVFS |
| | Probability: | |
| 43109, | 0.984 AA1: | |
| 43110 | 20 AA2: 21 | MKRRFAVLFFFTFVIPCVFS |
| | Probability: | |
| 9347, | 0.993 AA1: | |
| 9348 | 21 AA2: 22 | MKRRIAIIIIIITVLTVTAVA |
| | Probability: | |
| 72367, | 0.999 AA1: | |
| 72368 | 20 AA2: 21 | MKRRLSVLFLFTLCIPCVFS |
| | Probability: | |
| 30231, | 0.788 AA1: | |
| 30232 | 17 AA2: 18 | MKRRVVITGGAVISALG |
| | Probability: | |
| 12969, | 0.774 AA1: | |
| 12970 | 17 AA2: 18 | MKRRVVVTGGAVISALG |
| | Probability: | |
| 22771, | 1.000 AA1: | |
| 22772 | 27 AA2: 28 | MKRSLHRFLYILFAIAWTGFFASPLFA |
| | Probability: | |
| 19447, | 0.967 AA1: | |
| 19448 | 22 AA2: 23 | MKRSTRATFGRACRCTGVSAMS |
| | Probability: | |
| 68921, | 1.000 AA1: | |
| 68922 | 21 AA2: 22 | MKRSVLCLTALFVFFSVSAFS |

| | | |
|---|---|---|
| | Probability: | |
| 33589, | 1.000 AA1: | |
| 33590 | 26 AA2: 27 | MKRTPIISALLAILFIGLFTSCSTIS |
| | Probability: | |
| 74319, | 0.982 AA1: | |
| 74320 | 18 AA2: 19 | MKRVFFXALSCYTLGLNA |
| | Probability: | |
| 43167, | 0.999 AA1: | |
| 43168 | 21 AA2: 22 | MKRVSTGNFCAASLIASFAMS |
| | Probability: | |
| 573, | 1.000 AA1: | |
| 574 | 21 AA2: 22 | MKRVSTGNFCAASLMASFAMS |
| | Probability: | |
| 47261, | 0.994 AA1: | |
| 47262 | 27 AA2: 28 | MKRVVCMVTAAIMVAAVFTACKGKSSA |
| | Probability: | |
| 44543, | 0.581 AA1: | |
| 44544 | 16 AA2: 17 | MKRVVITGGSIVSALG |
| | Probability: | |
| 1193, | 1.000 AA1: | |
| 1194 | 19 AA2: 20 | MKRYAALIVIFLSFSAVYS |
| | Probability: | |
| 67719, | 0.999 AA1: | |
| 67720 | 19 AA2: 20 | MKRYAALTVFIFTFTAAYA |
| | Probability: | |
| 30135, | 0.995 AA1: | |
| 30136 | 18 AA2: 19 | MKRYLCLFAVLFTLLVFS |
| | Probability: | |
| 7171, | 1.000 AA1: | |
| 7172 | 20 AA2: 21 | MKRYLIVVCLAFFIAGFAAA |
| | Probability: | |
| 60659, | 0.762 AA1: | |
| 60660 | 20 AA2: 21 | MKRYVLDACALLAILGGENG |
| 61611, | Probability: | MKRYVLDACALLAVWG |

| | | |
|---|---|---|
| 61612 | 0.601 AA1: | |
| | 16 AA2: 17 | |
| | Probability: | |
| 51037, | 0.721 AA1: | |
| 51038 | 25 AA2: 26 | MKSCVSAIPPIAIPAAIIIAIPSFM |
| | Probability: | |
| 39997, | 0.966 AA1: | |
| 39998 | 19 AA2: 20 | MKSEFAKLAISATFGLALA |
| | Probability: | |
| 51569, | 0.763 AA1: | |
| 51570 | 19 AA2: 20 | MKSELLTAVCSCCIHPAGL |
| | Probability: | |
| 43361, | 0.999 AA1: | |
| 43362 | 21 AA2: 22 | MKSFCVFTVVFLFLGICSVAA |
| | Probability: | |
| 1261, | 0.713 AA1: | |
| 1262 | 29 AA2: 30 | MKSGLIVKSLGLETLPTAYILINSGPATA |
| | Probability: | |
| 71099, | 0.991 AA1: | |
| 71100 | 25 AA2: 26 | MKSIITLFVRRPVTVIMILAALIFA |
| | Probability: | |
| 22277, | 0.993 AA1: | |
| 22278 | 18 AA2: 19 | MKSIKFLPFLIASGFLAA |
| | Probability: | |
| 10817, | 1.000 AA1: | |
| 10818 | 24 AA2: 25 | MKSIKRTGYLLLTVVAIAAFSACA |
| | Probability: | |
| 31833, | 0.999 AA1: | |
| 31834 | 23 AA2: 24 | MKSKHGILFFFIGLILYCPSAFS |
| | Probability: | |
| 59807, | 0.617 AA1: | |
| 59808 | 23 AA2: 24 | MKSKIKTKLGSIIILHAILILFG |
| 43391, | Probability: | |
| 43392 | 1.000 AA1: | MKSKKIAMCFLAVAILCAVLNTNCSVRA |

| | | |
|---|---|---|
| | 28 AA2: 29 | |
| | Probability: | |
| 72849, | 1.000 AA1: | |
| 72850 | 21 AA2: 22 | MKSKLFLSVLLLLGAVFSIHA |
| | Probability: | |
| 64995, | 1.000 AA1: | |
| 64996 | 26 AA2: 27 | MKSKLTKFALSAAFGLALTFTLSCSG |
| | Probability: | |
| 62445, | 1.000 AA1: | |
| 62446 | 27 AA2: 28 | MKSKRKIFASFFLALALLLASSCGEFS |
| | Probability: | |
| 60223, | 0.833 AA1: | |
| 60224 | 25 AA2: 26 | MKSMKYFPLVQVFLFLFGLFLSVPA |
| | Probability: | |
| 55345, | 0.899 AA1: | |
| 55346 | 22 AA2: 23 | MKSNKYIFGIIVSMAIIFLTMA |
| | Probability: | |
| 32701, | 0.890 AA1: | |
| 32702 | 22 AA2: 23 | MKSNVLGTAVTIGGVVYGNVIA |
| | Probability: | |
| 39845, | 0.891 AA1: | |
| 39846 | 22 AA2: 23 | MKSNVLGTAVTIGGVVYGNVIA |
| | Probability: | |
| 33451, | 0.834 AA1: | |
| 33452 | 28 AA2: 29 | MKSPFLGAGFGVVSAFDFSKISSRPGRA |
| | Probability: | |
| 69459, | 0.915 AA1: | |
| 69460 | 21 AA2: 22 | MKSPGALICVGGVGSAFISDA |
| | Probability: | |
| 21887, | 0.847 AA1: | |
| 21888 | 21 AA2: 22 | MKSPITSILTAVTGSDASIMA |
| | Probability: | |
| 15111, | 0.949 AA1: | |
| 15112 | 28 AA2: 29 | MKSQIISVLISILXVFLIITFSNRSLIA |

| | | |
|---|---|---|
| | Probability: | |
| 51319, | 1.000 AA1: | |
| 51320 | 21 AA2: 22 | MKSRLLFTLCLLFTLTAAIYG |
| | Probability: | |
| 45231, | 1.000 AA1: | |
| 45232 | 25 AA2: 26 | MKSRRFLLFVLLTLSFSHFAPPLFA |
| | Probability: | |
| 56117, | 1.000 AA1: | |
| 56118 | 26 AA2: 27 | MKSSPSVKVFLITLGLLLYAPAVSQA |
| | Probability: | |
| 70595, | 1.000 AA1: | |
| 70596 | 26 AA2: 27 | MKSSSIRKFAALLSSAVLILISPAAV |
| | Probability: | |
| 64213, | 0.882 AA1: | |
| 64214 | 26 AA2: 27 | MKSSTHKWIVIAIFSATLFLSIIILT |
| | Probability: | |
| 59247, | 0.999 AA1: | |
| 59248 | 20 AA2: 21 | MKSTGILFLAAALFIFSCNT |
| | Probability: | |
| 58747, | 0.662 AA1: | |
| 58748 | 22 AA2: 23 | MKSTGTMPFASTYSLLMIGSRA |
| | Probability: | |
| 62959, | 0.999 AA1: | |
| 62960 | 26 AA2: 27 | MKSVKAFALIAALIVIVVFVGCKSSA |
| | Probability: | |
| 65113, | 0.988 AA1: | |
| 65114 | 20 AA2: 21 | MKSVKVFTLIAALMVIVVLA |
| | Probability: | |
| 61049, | 0.521 AA1: | |
| 61050 | 20 AA2: 21 | MKTALSIAGSDCSGGAGLQA |
| | Probability: | |
| 58261, | 0.951 AA1: | |
| 58262 | 26 AA2: 27 | MKTIKKIINDIMFVTGAALVILAAFS |
| 56809, | Probability: | MKTIKTIINFIMFVTGAAVVILALFS |

| | | |
|---|---|---|
| 56810 | 0.738 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 59045, | 0.807 AA1: | |
| 59046 | 26 AA2: 27 | MKTIKTIINYIMFVTGSALVILALFS |
| | Probability: | |
| 21507, | 0.985 AA1: | |
| 21508 | 21 AA2: 22 | MKTILKKNYIFVAAAILIVSA |
| | Probability: | |
| 41071, | 0.981 AA1: | |
| 41072 | 25 AA2: 26 | MKTILKKNYIFVAAAVLIVSAYFIG |
| | Probability: | |
| 39753, | 1.000 AA1: | |
| 39754 | 26 AA2: 27 | MKTIPSKLALAAGFLLAMALTFSCSS |
| | Probability: | |
| 62973, | 0.639 AA1: | |
| 62974 | 27 AA2: 28 | MKTKHNLLFNIAQTTVFTILLSVFLTA |
| | Probability: | |
| 71387, | 1.000 AA1: | |
| 71388 | 20 AA2: 21 | MKTKLSTAILFLAAALPALA |
| | Probability: | |
| 40371, | 1.000 AA1: | |
| 40372 | 25 AA2: 26 | MKTKNFLLATILGFALTFTFSGCVA |
| | Probability: | |
| 1523, | 0.910 AA1: | |
| 1524 | 25 AA2: 26 | MKTKPQLITVTIIAGLVLSCQFFLS |
| | Probability: | |
| 9033, | 1.000 AA1: | |
| 9034 | 23 AA2: 24 | MKTKTIKFYLIFLIIFSISPVFA |
| | Probability: | |
| 9973, | 0.997 AA1: | |
| 9974 | 25 AA2: 26 | MKTLFGKLLISFGVIILLMIVSVLA |
| 42717, | Probability: | |
| 42718 | 0.992 AA1: | MKTLFGKLLISFGVIILXMIVSVLA |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 61501, | 0.999 AA1: | |
| 61502 | 25 AA2: 26 | MKTLFGKLLLSFGIIIFMMISSVLA |
| | Probability: | |
| 37839, | 0.995 AA1: | |
| 37840 | 27 AA2: 28 | MKTLNRYRSASLFAAAAIAVCLSGCVK |
| | Probability: | |
| 46231, | 0.977 AA1: | |
| 46232 | 25 AA2: 26 | MKTLQLRRPLFIFICFILICKGIYA |
| | Probability: | |
| 28035, | 0.880 AA1: | |
| 28036 | 20 AA2: 21 | MKTLYLIFIFFVFCGILSFG |
| | Probability: | |
| 27545, | 0.986 AA1: | |
| 27546 | 24 AA2: 25 | MKTNISRNKKIYGICGVILLLILA |
| | Probability: | |
| 28789, | 0.997 AA1: | |
| 28790 | 22 AA2: 23 | MKTNKKFTVCRIAVFLAAVVFA |
| | Probability: | |
| 19543, | 1.000 AA1: | |
| 19544 | 24 AA2: 25 | MKTNRFLLTAALVLATALTLSCSS |
| | Probability: | |
| 401, | 0.998 AA1: | |
| 402 | 29 AA2: 30 | MKTRGYKLMYKKTVYLVFLVLICIGNIFA |
| | Probability: | |
| 51131, | 0.723 AA1: | |
| 51132 | 26 AA2: 27 | MKTRQIKLRLLIVFIVSLFYLMPLFL |
| | Probability: | |
| 37231, | 0.856 AA1: | |
| 37232 | 26 AA2: 27 | MKTRRFCIIYFAVTLLFLFYGCKSVS |
| | Probability: | |
| 71045, | 0.979 AA1: | |
| 71046 | 20 AA2: 21 | MKTYKFCAIFLAIFPLFSCA |

| | | |
|---|---|---|
| | Probability: | |
| 13243, | 0.918 AA1: | |
| 13244 | 20 AA2: 21 | MKVFVIAIILSLLFVGLLVG |
| | Probability: | |
| 47451, | 0.699 AA1: | |
| 47452 | 19 AA2: 20 | MKVIQTNAAPAAIGPYSQA |
| | Probability: | |
| 70759, | 0.998 AA1: | |
| 70760 | 22 AA2: 23 | MKVIRSKVLGFCMGVRRAVAMA |
| | Probability: | |
| 56815, | 0.603 AA1: | |
| 56816 | 19 AA2: 20 | MKVISTNSAPAAIGPYSQA |
| | Probability: | |
| 55189, | 0.691 AA1: | |
| 55190 | 29 AA2: 30 | MKVKKIKGVFIMKNVFKILTIIALAAVIV |
| | Probability: | |
| 62475, | 0.828 AA1: | |
| 62476 | 26 AA2: 27 | MKVKQSKTVPGTTKNPLFLLVLLFMA |
| | Probability: | |
| 28485, | 0.999 AA1: | |
| 28486 | 27 AA2: 28 | MKVLGMIKLGVILALFAAAACVMLAFV |
| | Probability: | |
| 73821, | 0.573 AA1: | |
| 73822 | 26 AA2: 27 | MKVSKLFSKHGGLFGAVAGEGSAVNN |
| | Probability: | |
| 25885, | 0.997 AA1: | |
| 25886 | 20 AA2: 21 | MKVTVASAALVVVVALEAAA |
| | Probability: | |
| 60459, | 0.558 AA1: | |
| 60460 | 25 AA2: 26 | MKVWLKLLIGSILGVIIGYLLPADN |
| | Probability: | |
| 24043, | 0.999 AA1: | |
| 24044 | 23 AA2: 24 | MKWLKIFAAIFLFLALTTCQTLS |
| 28701, | Probability: | MKWRFFMARKTGLTPKSLSSMLLDAGVVYA |

| | | |
|---|---|---|
| 28702 | 0.944 AA1: 30 AA2: 31 Probability: | |
| 33707, 33708 | 0.981 AA1: 21 AA2: 22 Probability: | MKXIVFIALALLICCGTLFFA |
| 66775, 66776 | 0.995 AA1: 19 AA2: 20 Probability: | MKYAWLGVLLVALGIQAYA |
| 34215, 34216 | 0.853 AA1: 25 AA2: 26 Probability: | MKYDVILAGVGGQGVLSVASIIAQA |
| 39879, 39880 | 0.829 AA1: 25 AA2: 26 Probability: | MKYDVILAGVGGQGVLSVSSIIALA |
| 34239, 34240 | 0.702 AA1: 21 AA2: 22 Probability: | MKYFNPANLAVIACPGGEIFA |
| 69859, 69860 | 0.962 AA1: 21 AA2: 22 Probability: | MKYFNSANLAIIACPGGETFA |
| 63829, 63830 | 0.605 AA1: 22 AA2: 23 Probability: | MKYGFMGPNYGTVSACASSSHA |
| 36771, 36772 | 0.999 AA1: 21 AA2: 22 Probability: | MKYILCLALXATFIFACASFK |
| 44789, 44790 | 0.996 AA1: 24 AA2: 25 Probability: | MKYILFALSLTILSCVSLGGRSIA |
| 36139, 36140 | 0.997 AA1: 21 AA2: 22 Probability: | MKYITCLALAITFIFSCASLK |
| 6035, 6036 | 0.781 AA1: | MKYKCWFVFFASLLIYA |

| | | |
|---|---|---|
| | 17 AA2: 18 | |
| | Probability: | |
| 11053, | 1.000 AA1: | |
| 11054 | 22 AA2: 23 | MKYKLLLPFLLFMFFCVSAGFT |
| | Probability: | |
| 58255, | 0.998 AA1: | |
| 58256 | 18 AA2: 19 | MKYLICAALAAAFISACA |
| | Probability: | |
| 63035, | 0.989 AA1: | |
| 63036 | 22 AA2: 23 | MKYLICPALAMALIFGCVSLQT |
| | Probability: | |
| 33633, | 0.981 AA1: | |
| 33634 | 18 AA2: 19 | MKYLIFPALAAAFILACA |
| | Probability: | |
| 69611, | 0.767 AA1: | |
| 69612 | 20 AA2: 21 | MKYLIYPALAAAFIFGCANL |
| | Probability: | |
| 66687, | 1.000 AA1: | |
| 66688 | 23 AA2: 24 | MKYLQMFLLVLVLMGIPISGTLA |
| | Probability: | |
| 50187, | 0.810 AA1: | |
| 50188 | 20 AA2: 21 | MKYNHKIPSACLXAWQPIFT |
| | Probability: | |
| 68317, | 0.755 AA1: | |
| 68318 | 20 AA2: 21 | MKYPLFIIVLSVFLFTQCKS |
| | Probability: | |
| 46957, | 0.991 AA1: | |
| 46958 | 21 AA2: 22 | MKYSNLANLAIIACPGGEVFA |
| | Probability: | |
| 48895, | 0.999 AA1: | |
| 48896 | 23 AA2: 24 | MKYVSKALFLFILFTGMAGIGHA |
| | Probability: | |
| 42103, | 0.925 AA1: | |
| 42104 | 25 AA2: 26 | MLAAAAVFGAAVMLLWNALXPQIFA |

| | | |
|---|---|---|
| | Probability: | |
| 35703, | 0.945 AA1: | |
| 35704 | 22 AA2: 23 | MLAALSIPVSAAAGICVLSVMG |
| | Probability: | |
| 7073, | 0.858 AA1: | |
| 7074 | 19 AA2: 20 | MLAGMILLFFAPFAPPAQQ |
| | Probability: | |
| 18839, | 0.927 AA1: | |
| 18840 | 16 AA2: 17 | MLAGVFCVSFAAGSLA |
| | Probability: | |
| 30583, | 0.969 AA1: | |
| 30584 | 19 AA2: 20 | MLAIKSVLSGCGTSTGSTA |
| | Probability: | |
| 24747, | 0.848 AA1: | |
| 24748 | 19 AA2: 20 | MLAISGGIVAIATTLLLVA |
| | Probability: | |
| 49179, | 0.992 AA1: | |
| 49180 | 24 AA2: 25 | MLAIVVLSLIVSIMLALAVTRSIA |
| | Probability: | |
| 6997, | 0.786 AA1: | |
| 6998 | 26 AA2: 27 | MLARRFSIIGIFSVTWVIRSLCMALA |
| | Probability: | |
| 35611, | 0.905 AA1: | |
| 35612 | 23 AA2: 24 | MLASSRAASLLGPMIASPAPRNV |
| | Probability: | |
| 13621, | 0.993 AA1: | |
| 13622 | 23 AA2: 24 | MLAVIAAMVPALAASVVFFGVEA |
| | Probability: | |
| 51877, | 0.726 AA1: | |
| 51878 | 15 AA2: 16 | MLAVLAALAAGAVWG |
| | Probability: | |
| 7339, | 0.983 AA1: | |
| 7340 | 23 AA2: 24 | MLCSFPVSLSFALTCTMPFASIS |
| 4711, | Probability: | MLEKLPGHNIVFFLLVLAMLAFYSSTFDA |

| | | |
|---|---|---|
| 4712 | 0.893 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| 10687, | 0.991 AA1: | |
| 10688 | 21 AA2: 22 | MLESACMAATAGGFATVVCMA |
| | Probability: | |
| 16887, | 0.528 AA1: | |
| 16888 | 22 AA2: 23 | MLETFFIAVTASPVLTCMPSMR |
| | Probability: | |
| 15775, | 0.941 AA1: | |
| 15776 | 15 AA2: 16 | MLFAALALAAHIAIA |
| | Probability: | |
| 21715, | 0.948 AA1: | |
| 21716 | 25 AA2: 26 | MLFAILLVLFLIVCVFICLLILVQS |
| | Probability: | |
| 72047, | 0.862 AA1: | |
| 72048 | 28 AA2: 29 | MLFAILLVLFLIVCVLICLLILIQSDKG |
| | Probability: | |
| 41129, | 0.676 AA1: | |
| 41130 | 16 AA2: 17 | MLFFLMFIGGCSLSTA |
| | Probability: | |
| 44461, | 0.660 AA1: | |
| 44462 | 23 AA2: 24 | MLFFYSSSASSLIIVLYSLPESK |
| | Probability: | |
| 69809, | 0.994 AA1: | |
| 69810 | 22 AA2: 23 | MLFGILIICRLAALMLILPVLT |
| | Probability: | |
| 71655, | 0.998 AA1: | |
| 71656 | 19 AA2: 20 | MLFILKMWASVISGSIALA |
| | Probability: | |
| 32635, | 0.763 AA1: | |
| 32636 | 29 AA2: 30 | MLFIMIKLMRKTIGALLPCVAITILVING |
| 6345, | Probability: | |
| 6346 | 0.509 AA1: | MLFIPKNLRLLVLFYILIAGVLML |

| | | |
|---|---|---|
| | 24 AA2: 25 | |
| | Probability: | |
| 59763, | 0.597 AA1: | |
| 59764 | 19 AA2: 20 | MLFLIILVVGGLAFFFSMR |
| | Probability: | |
| 73003, | 0.782 AA1: | |
| 73004 | 23 AA2: 24 | MLFLSVSVITFGLALCGAGCLFS |
| | Probability: | |
| 30081, | 0.949 AA1: | |
| 30082 | 28 AA2: 29 | MLFNRKGRHIMNIRSKIMQSIAVGAAVA |
| | Probability: | |
| 21997, | 0.971 AA1: | |
| 21998 | 28 AA2: 29 | MLFPWWKRFLPDISAALLLLFSMGIGMT |
| | Probability: | |
| 77393, | 0.524 AA1: | |
| 77394 | 29 AA2: 30 | MLFQVLLPIGLWAQFALNIPVWLFPVALA |
| | Probability: | |
| 18259, | 0.863 AA1: | |
| 18260 | 28 AA2: 29 | MLFRDESPAVCAMVFSRLPSKLSAAALA |
| | Probability: | |
| 55151, | 0.602 AA1: | |
| 55152 | 20 AA2: 21 | MLFSLGSGGLWSRVFPSVLA |
| | Probability: | |
| 77195, | 1.000 AA1: | |
| 77196 | 21 AA2: 22 | MLFSSCLRATCAAFAPSSATA |
| | Probability: | |
| 5265, | 0.740 AA1: | |
| 5266 | 31 AA2: 32 | MLFWGFSFISIKITVAVFPPMTLGALRFALA |
| | Probability: | |
| 25321, | 0.982 AA1: | |
| 25322 | 19 AA2: 20 | MLGATTTTTAGATATXVAA |
| | Probability: | |
| 51323, | 0.846 AA1: | |
| 51324 | 21 AA2: 22 | MLGFVIAAFIAGIGGCLYAMV |

| | | |
|---|---|---|
| | Probability: | |
| 9745, | 0.807 AA1: | |
| 9746 | 21 AA2: 22 | MLGFVTAAFIAGIGGCLYAMV |
| | Probability: | |
| 14387, | 0.988 AA1: | |
| 14388 | 28 AA2: 29 | MLHAFPGLLLLILLQLFASVSGLFPVTA |
| | Probability: | |
| 37597, | 0.964 AA1: | |
| 37598 | 27 AA2: 28 | MLHIWMKRFIIPLILGAAFFFSCITNS |
| | Probability: | |
| 56353, | 0.907 AA1: | |
| 56354 | 19 AA2: 20 | MLHNFLIFISPFYFPALLA |
| | Probability: | |
| 9701, | 0.689 AA1: | |
| 9702 | 22 AA2: 23 | MLIASRLAFSLTRAGWSLVATT |
| | Probability: | |
| 28299, | 0.856 AA1: | |
| 28300 | 21 AA2: 22 | MLIELSGAVTTVSGTARVTSA |
| | Probability: | |
| 71729, | 0.929 AA1: | |
| 71730 | 25 AA2: 26 | MLIFSMGLLFAIFAAALLFMYKKNA |
| | Probability: | |
| 14273, | 0.591 AA1: | |
| 14274 | 19 AA2: 20 | MLIGMLINMMFMITSMGLA |
| | Probability: | |
| 13751, | 0.526 AA1: | |
| 13752 | 25 AA2: 26 | MLIGVVVTSGSGGPSVAGLWFSANA |
| | Probability: | |
| 58707, | 0.821 AA1: | |
| 58708 | 23 AA2: 24 | MLINVKDAVVGLLVAVFLGALVA |
| | Probability: | |
| 21691, | 0.981 AA1: | |
| 21692 | 22 AA2: 23 | MLIRSGLSLGNVLFSLGSAQDA |
| 50657, | Probability: | MLISAMTAFMGLYLMDSFHVLK |

| | | |
|---|---|---|
| 50658 | 0.564 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 6801, | 0.762 AA1: | |
| 6802 | 27 AA2: 28 | MLISLLIPIISVGGLALLFGIILGFSA |
| | Probability: | |
| 48367, | 0.888 AA1: | |
| 48368 | 21 AA2: 22 | MLISYFDSSFLLAILFNESHA |
| | Probability: | |
| 2727, | 0.626 AA1: | |
| 2728 | 18 AA2: 19 | MLITVISFALGVYFSTKA |
| | Probability: | |
| 32335, | 0.636 AA1: | |
| 32336 | 18 AA2: 19 | MLITVISFALGVYFSTKA |
| | Probability: | |
| 51195, | 0.909 AA1: | |
| 51196 | 19 AA2: 20 | MLITVLGLVLILLARSRTA |
| | Probability: | |
| 45061, | 0.999 AA1: | |
| 45062 | 21 AA2: 22 | MLIWAICVLLAIFLFTPIALA |
| | Probability: | |
| 67473, | 0.543 AA1: | |
| 67474 | 27 AA2: 28 | MLKALLPSSWAAAAVGPKICIPCFLNS |
| | Probability: | |
| 939, | 0.618 AA1: | |
| 940 | 32 AA2: 33 | MLKALLPSSWAAAAVGPKICIPCFLNSSTTPA |
| | Probability: | |
| 61755, | 0.546 AA1: | |
| 61756 | 26 AA2: 27 | MLKDISFEIKKGSVAALVGASGSGKS |
| | Probability: | |
| 12351, | 0.982 AA1: | |
| 12352 | 23 AA2: 24 | MLKFLELTGIFLLFFSAAGPSMK |
| 5931, | Probability: | |
| 5932 | 0.935 AA1: | MLKISAILIGIIAGYILAIILGQVNFA |

| | | |
|---|---|---|
| | 27 AA2: 28 | |
| | Probability: | |
| 3127, | 0.968 AA1: | |
| 3128 | 19 AA2: 20 | MLKLALGYGGGVGAFSTFA |
| | Probability: | |
| 57795, | 1.000 AA1: | |
| 57796 | 27 AA2: 28 | MLKLKFIGLLSCLLLLALGACGSSPQS |
| | Probability: | |
| 74695, | 0.851 AA1: | |
| 74696 | 21 AA2: 22 | MLKPNKPLIVIFLGPAVICYA |
| | Probability: | |
| 32511, | 0.999 AA1: | |
| 32512 | 29 AA2: 30 | MLKSKIKILLFVFLMTALVLYTSCSGKKA |
| | Probability: | |
| 639, | 0.998 AA1: | |
| 640 | 26 AA2: 27 | MLLLRKTQICMGFIALQALFAGRLVA |
| | Probability: | |
| 3143, | 0.998 AA1: | |
| 3144 | 26 AA2: 27 | MLLMKKHTLFWGCCLVSLIAATTVCA |
| | Probability: | |
| 76491, | 1.000 AA1: | |
| 76492 | 19 AA2: 20 | MLLQKKILVILLLACAAQA |
| | Probability: | |
| 66293, | 1.000 AA1: | |
| 66294 | 20 AA2: 21 | MLLRLLAAVCALAASGVIMA |
| | Probability: | |
| 1597, | 0.988 AA1: | |
| 1598 | 31 AA2: 32 | MLLSNKIAKASRAFSTNSTLIALVMSLSSSS |
| | Probability: | |
| 5073, | 0.718 AA1: | |
| 5074 | 21 AA2: 22 | MLLSVFVITIAGSFLALFLLS |
| | Probability: | |
| 48091, | 0.674 AA1: | |
| 48092 | 27 AA2: 28 | MLLTIFFIMNKVLIIGAGGVANVVAQK |

| | | |
|---|---|---|
| | Probability: | |
| 58779, | 0.657 AA1: | |
| 58780 | 24 AA2: 25 | MLLTPNPKGSTTMRKLILTTAAVA |
| | Probability: | |
| 68777, | 0.964 AA1: | |
| 68778 | 28 AA2: 29 | MLLTPNPKGSTTMRKLILTTAAVAAAIA |
| | Probability: | |
| 10683, | 0.571 AA1: | |
| 10684 | 16 AA2: 17 | MLLVTLSAFVFPYIFA |
| | Probability: | |
| 74759, | 0.527 AA1: | |
| 74760 | 17 AA2: 18 | MLMFLFGWRMRSSAAAA |
| | Probability: | |
| 41339, | 0.602 AA1: | |
| 41340 | 28 AA2: 29 | MLMLAAIPFSLFASITTLLLLFTRSVPT |
| | Probability: | |
| 56491, | 0.989 AA1: | |
| 56492 | 20 AA2: 21 | MLMLLVGFLVGGAAASIVWA |
| | Probability: | |
| 68109, | 0.983 AA1: | |
| 68110 | 26 AA2: 27 | MLMNKYLIYSFLIIATVIFAITACNA |
| | Probability: | |
| 64351, | 1.000 AA1: | |
| 64352 | 23 AA2: 24 | MLMNRQLWLIFFSLLLIASFAHA |
| | Probability: | |
| 31179, | 0.562 AA1: | |
| 31180 | 27 AA2: 28 | MLNEVIRLQGKICGLMLAMGFSQKTEA |
| | Probability: | |
| 74827, | 0.616 AA1: | |
| 74828 | 18 AA2: 19 | MLNGSFGTSLSKTGSASG |
| | Probability: | |
| 55687, | 0.965 AA1: | |
| 55688 | 23 AA2: 24 | MLNKIVLIFSLCFSLVFSGCTRR |
| 38221, | Probability: | MLNLRVSAIAMPPFAVPSSFV |

| | | |
|---|---|---|
| 38222 | 0.595 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| 51295, | 0.766 AA1: | |
| 51296 | 26 AA2: 27 | MLNQKKDPILIVKFCLLVLIAAFAFS |
| | Probability: | |
| 33535, | 0.809 AA1: | |
| 33536 | 29 AA2: 30 | MLNQKKDPILIVKVCLLILIGAFAFSFIS |
| | Probability: | |
| 73175, | 0.992 AA1: | |
| 73176 | 27 AA2: 28 | MLPALLLMVFAPALMGQSNGVGGEAQA |
| | Probability: | |
| 36789, | 0.944 AA1: | |
| 36790 | 22 AA2: 23 | MLPFLILLLCLLVGIIAVTSFK |
| | Probability: | |
| 7853, | 0.946 AA1: | |
| 7854 | 16 AA2: 17 | MLPGFIFSVGLSPALA |
| | Probability: | |
| 35499, | 0.614 AA1: | |
| 35500 | 22 AA2: 23 | MLPSALPGIVTSVILAIGRVVG |
| | Probability: | |
| 11897, | 0.856 AA1: | |
| 11898 | 19 AA2: 20 | MLPVSCSCIALRMLSSCSA |
| | Probability: | |
| 40411, | 0.974 AA1: | |
| 40412 | 29 AA2: 30 | MLQKNPSLDPAKISLPAVIAGLTAGLSCA |
| | Probability: | |
| 6205, | 0.778 AA1: | |
| 6206 | 20 AA2: 21 | MLQLFTLLLLAVALVGCGQK |
| | Probability: | |
| 20307, | 0.523 AA1: | |
| 20308 | 21 AA2: 22 | MLQNTNLSSALFTDFYSLTMA |
| 72937, | Probability: | |
| 72938 | 0.973 AA1: | MLQNYSCAGSPFLMFLPFICLA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 36591, | 0.894 AA1: | |
| 36592 | 21 AA2: 22 | MLQRAFNFALLLQILSKIALA |
| | Probability: | |
| 6151, | 0.926 AA1: | |
| 6152 | 29 AA2: 30 | MLQYPQRCEMKKQVIFILTALIAFCACSA |
| | Probability: | |
| 6197, | 0.965 AA1: | |
| 6198 | 18 AA2: 19 | MLRFKFILIVSISVLVFG |
| | Probability: | |
| 62123, | 1.000 AA1: | |
| 62124 | 24 AA2: 25 | MLRFLILSILLLAAFLLTVFHGSA |
| | Probability: | |
| 51605, | 0.972 AA1: | |
| 51606 | 29 AA2: 30 | MLRKFRIRTRLLFSFFIVVLFTLIVGLTG |
| | Probability: | |
| 73013, | 1.000 AA1: | |
| 73014 | 23 AA2: 24 | MLRKVLLFTLAVFFLSGSWLAAA |
| | Probability: | |
| 25131, | 0.522 AA1: | |
| 25132 | 20 AA2: 21 | MLRLGAKLFARFLILGLSMA |
| | Probability: | |
| 28505, | 1.000 AA1: | |
| 28506 | 21 AA2: 22 | MLRSIFLPAFLIALSLQNAFA |
| | Probability: | |
| 22987, | 0.675 AA1: | |
| 22988 | 26 AA2: 27 | MLRVSRLALRACALFFPSLPPENSLN |
| | Probability: | |
| 41425, | 1.000 AA1: | |
| 41426 | 22 AA2: 23 | MLSASFRTSALSLSLLSSSAWS |
| | Probability: | |
| 59971, | 1.000 AA1: | |
| 59972 | 22 AA2: 23 | MLSASFRTSALSLSLLSSSAWS |

| | | |
|---|---|---|
| | Probability: | |
| 24809, | 0.997 AA1: | |
| 24810 | 23 AA2: 24 | MLSASFRTSALSLSRLSSSACSS |
| | Probability: | |
| 73763, | 0.523 AA1: | |
| 73764 | 28 AA2: 29 | MLSCCQTRRRVMFFYYAFAVSATGAGLA |
| | Probability: | |
| 58015, | 0.982 AA1: | |
| 58016 | 26 AA2: 27 | MLSKFXDVMIGIVILLVFAASMSTLS |
| | Probability: | |
| 17073, | 1.000 AA1: | |
| 17074 | 21 AA2: 22 | MLSLNKNVFLILCLTAVQAFA |
| | Probability: | |
| 16739, | 0.571 AA1: | |
| 16740 | 20 AA2: 21 | MLSMPIMFSAMSYGSISYNA |
| | Probability: | |
| 56489, | 1.000 AA1: | |
| 56490 | 25 AA2: 26 | MLSMRFAMSSALSALWFASSLISFA |
| | Probability: | |
| 3575, | 1.000 AA1: | |
| 3576 | 28 AA2: 29 | MLSMRSAMSSALCALWFASSFISPATTA |
| | Probability: | |
| 44535, | 0.990 AA1: | |
| 44536 | 27 AA2: 28 | MLSNPDFTSAVAMASAINGVFPNTAIA |
| | Probability: | |
| 43541, | 0.765 AA1: | |
| 43542 | 20 AA2: 21 | MLSPVIIATSMPILCRLSIA |
| | Probability: | |
| 49653, | 0.555 AA1: | |
| 49654 | 24 AA2: 25 | MLSRPSSPLMGYPGALGIDLSAEA |
| | Probability: | |
| 52955, | 0.895 AA1: | |
| 52956 | 23 AA2: 24 | MLSSELKSGKSSMPLLVSSGVAA |
| 44921, | Probability: | MLSVMNKRFKILDIVIILLAVCLTVSSA |

| | | |
|---|---|---|
| 44922 | 0.989 AA1: | |
| | 28 AA2: 29 | |
| | Probability: | |
| 57315, | 0.979 AA1: | |
| 57316 | 25 AA2: 26 | MLSVVAAVKTILLYSIALLVSISFA |
| | Probability: | |
| 23729, | 0.626 AA1: | |
| 23730 | 28 AA2: 29 | MLTFAAIAFSLVASITTLLLLLIRTVPT |
| | Probability: | |
| 64533, | 0.924 AA1: | |
| 64534 | 25 AA2: 26 | MLTHRSRPPAVSAIVLNIASTEVSS |
| | Probability: | |
| 1673, | 0.892 AA1: | |
| 1674 | 20 AA2: 21 | MLTKTLSIGGMTCAACSARV |
| | Probability: | |
| 68699, | 0.989 AA1: | |
| 68700 | 22 AA2: 23 | MLTMIKTMNKRITVALSALVTA |
| | Probability: | |
| 32573, | 0.781 AA1: | |
| 32574 | 20 AA2: 21 | MLTRCAKVMASGLSSGVSFM |
| | Probability: | |
| 28591, | 0.509 AA1: | |
| 28592 | 20 AA2: 21 | MLTRSAKVMASGLPGGVSFM |
| | Probability: | |
| 21815, | 0.522 AA1: | |
| 21816 | 27 AA2: 28 | MLTSITIPNSVTSIVNGAFLGCTSLTA |
| | Probability: | |
| 76863, | 0.998 AA1: | |
| 76864 | 24 AA2: 25 | MLTSSSTATSSPSCSTASPTATTA |
| | Probability: | |
| 69821, | 0.648 AA1: | |
| 69822 | 18 AA2: 19 | MLVKRAFILIFLVLSCCG |
| 73431, | Probability: | |
| 73432 | 0.937 AA1: | MLVLHIILIVAAMLTSTRAGFG |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 52875, | 0.708 AA1: | |
| 52876 | 24 AA2: 25 | MLVMVSFHTPMLLAMAVMGIVFFA |
| | Probability: | |
| 28917, | 0.738 AA1: | |
| 28918 | 21 AA2: 22 | MLVPGTAATPAPAATPAPATA |
| | Probability: | |
| 76905, | 1.000 AA1: | |
| 76906 | 20 AA2: 21 | MLVPLLFLLLLMFLSLPVMS |
| | Probability: | |
| 78331, | 0.998 AA1: | |
| 78332 | 22 AA2: 23 | MLVSLLLILPFVFAFLMLASSS |
| | Probability: | |
| 42097, | 0.559 AA1: | |
| 42098 | 28 AA2: 29 | MLVSPVSRYSIVIGKILGSGFAAIISMA |
| | Probability: | |
| 73933, | 0.950 AA1: | |
| 73934 | 23 AA2: 24 | MLWLPVLGLCVLPPAWAGMLLLA |
| | Probability: | |
| 57227, | 1.000 AA1: | |
| 57228 | 26 AA2: 27 | MLWRKKGMKKLLAVLLVLVVASAAFA |
| | Probability: | |
| 24433, | 0.909 AA1: | |
| 24434 | 29 AA2: 30 | MLYLLNGEFMKKAFFFLLLFFSFYGVCFS |
| | Probability: | |
| 75763, | 0.547 AA1: | |
| 75764 | 24 AA2: 25 | MLYVFAMGIAVSIINIASMSISMG |
| | Probability: | |
| 65391, | 0.958 AA1: | |
| 65392 | 25 AA2: 26 | MLYYYLCAVFTLISSAVSFGFSINA |
| | Probability: | |
| 55935, | 0.975 AA1: | |
| 55936 | 19 AA2: 20 | MMAEAALWLGALLTGEAQS |

| | | |
|---|---|---|
| | Probability: | |
| 19607, | 0.785 AA1: | |
| 19608 | 18 AA2: 19 | MMAEAYSIYFFLAFLALA |
| | Probability: | |
| 5499, | 0.791 AA1: | |
| 5500 | 20 AA2: 21 | MMAECALALVSTSLLGMVDA |
| | Probability: | |
| 50079, | 0.964 AA1: | |
| 50080 | 23 AA2: 24 | MMAEFVCLLVLMLVPACSKKTSA |
| | Probability: | |
| 12459, | 0.985 AA1: | |
| 12460 | 19 AA2: 20 | MMAEIGLACAMVLVMAAFG |
| | Probability: | |
| 31073, | 0.883 AA1: | |
| 31074 | 17 AA2: 18 | MMAEMLATILRSPQATA |
| | Probability: | |
| 9697, | 0.600 AA1: | |
| 9698 | 32 AA2: 33 | MMAEYRFMRTLVQNIRNGLMASSKLAAQLVTA |
| | Probability: | |
| 75581, | 1.000 AA1: | |
| 75582 | 29 AA2: 30 | MMALAALGLLLAAAACSSTGGAFPAPPQA |
| | Probability: | |
| 60503, | 0.880 AA1: | |
| 60504 | 18 AA2: 19 | MMALFVAILVTSNVASSA |
| | Probability: | |
| 58109, | 0.992 AA1: | |
| 58110 | 28 AA2: 29 | MMANSVPCGFIAACLFATCPTSSLPPSA |
| | Probability: | |
| 45479, | 0.604 AA1: | |
| 45480 | 27 AA2: 28 | MMDFFTKQAYAIGRLCLLIAILSAALN |
| | Probability: | |
| 10845, | 0.771 AA1: | |
| 10846 | 26 AA2: 27 | MMGINKLYKSFIMTAVLCLLPLLLAS |
| 60003, | Probability: | MMGRVLLALAPASVFGVILFG |

| | | |
|---|---|---|
| 60004 | 0.887 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| 58049, | 0.789 AA1: | |
| 58050 | 18 AA2: 19 | MMGTILMSSSSTAGSTSG |
| | Probability: | |
| 16657, | 0.942 AA1: | |
| 16658 | 25 AA2: 26 | MMHAAQTYFFITALSFPFLALYTAA |
| | Probability: | |
| 5821, | 1.000 AA1: | |
| 5822 | 19 AA2: 20 | MMIAVLCLVIGLGMGSAMA |
| | Probability: | |
| 18209, | 0.744 AA1: | |
| 18210 | 23 AA2: 24 | MMIIIPSVLPGTITAIFCGYAFR |
| | Probability: | |
| 59757, | 0.999 AA1: | |
| 59758 | 27 AA2: 28 | MMIKPFSSLSSCLFVVLFSLLTASLTG |
| | Probability: | |
| 3917, | 0.631 AA1: | |
| 3918 | 26 AA2: 27 | MMKCDVILAGVGGQGVLSVAAIIAQA |
| | Probability: | |
| 33135, | 0.976 AA1: | |
| 33136 | 21 AA2: 22 | MMKIGIKFTIIMVVLGLFAIA |
| | Probability: | |
| 56451, | 0.967 AA1: | |
| 56452 | 25 AA2: 26 | MMKIIKGIMQYAPVGVCALVFGVVA |
| | Probability: | |
| 777, | 0.968 AA1: | |
| 778 | 27 AA2: 28 | MMKKERFILTVTLCLLPLLFASCKVYT |
| | Probability: | |
| 76987, | 0.996 AA1: | |
| 76988 | 24 AA2: 25 | MMKKIRSNRLFPLLLLFTPLLSFV |
| 29417, | Probability: | |
| 29418 | 0.999 AA1: | MMKKLFFTFFLIFTLFLPLSA |

| | | |
|---|---|---|
| | 21 AA2: 22 | |
| | Probability: | |
| 61099, | 0.996 AA1: | |
| 61100 | 20 AA2: 21 | MMKKPVVIGLAVVVLAAVVA |
| | Probability: | |
| 52555, | 0.509 AA1: | |
| 52556 | 20 AA2: 21 | MMKKRIAVLGATGSIGKSSL |
| | Probability: | |
| 68755, | 0.509 AA1: | |
| 68756 | 20 AA2: 21 | MMKKRIAVLGATGSIGKSSL |
| | Probability: | |
| 24465, | 0.779 AA1: | |
| 24466 | 19 AA2: 20 | MMKKTIACIGSGNMGPALM |
| | Probability: | |
| 13889, | 0.996 AA1: | |
| 13890 | 21 AA2: 22 | MMKMGVMGAALSTXLSRIIAA |
| | Probability: | |
| 21673, | 1.000 AA1: | |
| 21674 | 19 AA2: 20 | MMKPLFFSVLILFSSGLFA |
| | Probability: | |
| 67067, | 1.000 AA1: | |
| 67068 | 19 AA2: 20 | MMKPLFISVFILVSAGLFA |
| | Probability: | |
| 47723, | 1.000 AA1: | |
| 47724 | 21 AA2: 22 | MMKRLIFVLALVFIASAFIFA |
| | Probability: | |
| 31609, | 1.000 AA1: | |
| 31610 | 21 AA2: 22 | MMKRLVFVLALVFIASAFIFA |
| | Probability: | |
| 13351, | 0.998 AA1: | |
| 13352 | 25 AA2: 26 | MMKSVSYLRFLAVVIIFLLLVPIFA |
| | Probability: | |
| 71071, | 0.968 AA1: | |
| 71072 | 25 AA2: 26 | MMKTQKIKSKAEMVAVACAVIVAVA |

| | | |
|---|---|---|
| | Probability: | |
| 43981, | 0.723 AA1: | |
| 43982 | 25 AA2: 26 | MMKTSLLAGLTGMTGAPPRRPPNSA |
| | Probability: | |
| 73615, | 0.688 AA1: | |
| 73616 | 27 AA2: 28 | MMLGAEMKTRFYLPVFLGAAIAAVFLA |
| | Probability: | |
| 63403, | 1.000 AA1: | |
| 63404 | 25 AA2: 26 | MMLKKINLLCLLLCLSFLALGACKS |
| | Probability: | |
| 14873, | 0.959 AA1: | |
| 14874 | 29 AA2: 30 | MMLYSKDKNKPIKLTAVSLFSGCGGFDWG |
| | Probability: | |
| 71383, | 0.542 AA1: | |
| 71384 | 22 AA2: 23 | MMNAVNNIAAALAATGSGKSEG |
| | Probability: | |
| 64653, | 0.612 AA1: | |
| 64654 | 26 AA2: 27 | MMNKRKISNINLHGLCLIIFLLFSVK |
| | Probability: | |
| 23899, | 0.993 AA1: | |
| 23900 | 27 AA2: 28 | MMNLGLLGAGLVMGIAAVGSGIGVGFA |
| | Probability: | |
| 51271, | 0.977 AA1: | |
| 51272 | 27 AA2: 28 | MMNLGLLGPGLVMGIAAVGSAIGIGFA |
| | Probability: | |
| 56255, | 0.966 AA1: | |
| 56256 | 27 AA2: 28 | MMNLGLLGPGLVMGIAAVGSAIGIGFA |
| | Probability: | |
| 8449, | 0.911 AA1: | |
| 8450 | 22 AA2: 23 | MMNQFIGTVSITGGTVSATGYA |
| | Probability: | |
| 6869, | 0.605 AA1: | |
| 6870 | 23 AA2: 24 | MMNRGLAIGMASALMLTMIRILL |
| 73157, | Probability: | MMPGMPAAGCRGPAPAAGPRAA |

| | | |
|---|---|---|
| 73158 | 0.640 AA1: 22 AA2: 23 Probability: | |
| 61819, 61820 | 0.972 AA1: 26 AA2: 27 Probability: | MMPRHIYYAKVRYMAMILCALGIAFA |
| 29461, 29462 | 0.997 AA1: 27 AA2: 28 Probability: | MMPRTSVISISFSARSAAAIFAAAVSA |
| 22873, 22874 | 0.754 AA1: 22 AA2: 23 Probability: | MMPSGVISITPISVMIMSTASK |
| 48503, 48504 | 0.601 AA1: 16 AA2: 17 Probability: | MMRAAAFLCANTVIHA |
| 40901, 40902 | 0.844 AA1: 26 AA2: 27 Probability: | MMRFLPRDVCAYLGMTGGITVLSSQA |
| 8447, 8448 | 1.000 AA1: 21 AA2: 22 Probability: | MMRFVFSAVILILLGSFPAAA |
| 32977, 32978 | 1.000 AA1: 21 AA2: 22 Probability: | MMRFVFSAVILILLGSFSAFA |
| 1271, 1272 | 1.000 AA1: 23 AA2: 24 Probability: | MMRGFKVAVTVLFAAAAFSAASA |
| 14097, 14098 | 0.676 AA1: 22 AA2: 23 Probability: | MMRGIVCGWVVSGTGSVSGTHV |
| 17939, 17940 | 0.995 AA1: 20 AA2: 21 Probability: | MMRITVAAVLTVIAAVNLNA |
| 78621, 78622 | 0.988 AA1: | MMRLAYRTTLPAGLFLAGFFLAGCFG |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 2305, | 1.000 AA1: | |
| 2306 | 29 AA2: 30 | MMRLIKMLNCKLMTLVFFIVFSILFPAVA |
| | Probability: | |
| 1761, | 0.691 AA1: | |
| 1762 | 17 AA2: 18 | MMRMCAFRSVGMTVSQA |
| | Probability: | |
| 45043, | 0.991 AA1: | |
| 45044 | 26 AA2: 27 | MMRNKFAPLFLALLLCSCRSINHSRA |
| | Probability: | |
| 77967, | 0.940 AA1: | |
| 77968 | 21 AA2: 22 | MMSATVLWVIGGVLCALVELA |
| | Probability: | |
| 42463, | 0.949 AA1: | |
| 42464 | 18 AA2: 19 | MMSAVMMLLALAVLALFV |
| | Probability: | |
| 57027, | 0.960 AA1: | |
| 57028 | 21 AA2: 22 | MMSIISCIGSGNMGFALMSAA |
| | Probability: | |
| 56111, | 0.882 AA1: | |
| 56112 | 29 AA2: 30 | MMSKGAKMDTGLFVATLLLLGLGIVLVYS |
| | Probability: | |
| 70467, | 0.513 AA1: | |
| 70468 | 26 AA2: 27 | MMSLLRITKTVFKKYAQVFAVFLAFA |
| | Probability: | |
| 40501, | 0.993 AA1: | |
| 40502 | 24 AA2: 25 | MMSNKSVFKSAAAVILATLFLAIA |
| | Probability: | |
| 13287, | 1.000 AA1: | |
| 13288 | 21 AA2: 22 | MMSRKGIIIFVFLLLAQGVFA |
| | Probability: | |
| 43229, | 0.776 AA1: | |
| 43230 | 21 AA2: 22 | MMTSLILPASSFETPSLSGFG |

| | | |
|---|---|---|
| | Probability: | |
| 27065, | 0.852 AA1: | |
| 27066 | 20 AA2: 21 | MMVQNLTLSILLLFIMSCDS |
| | Probability: | |
| 23189, | 0.900 AA1: | |
| 23190 | 26 AA2: 27 | MMVTGAGPVVKGVLLALLAFSLFSWA |
| | Probability: | |
| 65271, | 0.694 AA1: | |
| 65272 | 26 AA2: 27 | MNAIRKAATNKIIRTISMIIIAAVLS |
| | Probability: | |
| 17791, | 0.922 AA1: | |
| 17792 | 25 AA2: 26 | MNAIRLISHLLISLMLFTSCGKNNS |
| | Probability: | |
| 56345, | 0.926 AA1: | |
| 56346 | 27 AA2: 28 | MNASRYIKIALFFIVLGSAGAGYIVLS |
| | Probability: | |
| 4403, | 0.999 AA1: | |
| 4404 | 21 AA2: 22 | MNAVKTIALIAAVAWAAPTQS |
| | Probability: | |
| 75333, | 0.857 AA1: | |
| 75334 | 25 AA2: 26 | MNCKYSRMLVPILAALLLLLGGCGT |
| | Probability: | |
| 63945, | 0.667 AA1: | |
| 63946 | 21 AA2: 22 | MNDRPYVSSSFLSTAISNVFG |
| | Probability: | |
| 26251, | 0.565 AA1: | |
| 26252 | 25 AA2: 26 | MNERYRFLFSGSGGQGVISMGILLA |
| | Probability: | |
| 76237, | 0.975 AA1: | |
| 76238 | 20 AA2: 21 | MNESVLWAAILVVMGLSLLA |
| | Probability: | |
| 67895, | 0.983 AA1: | |
| 67896 | 21 AA2: 22 | MNFAKKILCIGAGYVGGPTMA |
| 77917, | Probability: | MNFDYKKPLGKVFIVFLFFFILAHSALYA |

| | | |
|---|---|---|
| 77918 | 0.982 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| 42193, | 1.000 AA1: | |
| 42194 | 25 AA2: 26 | MNFFPRVKIFLIILGFSFAPAVSQA |
| | Probability: | |
| 75369, | 0.997 AA1: | |
| 75370 | 19 AA2: 20 | MNFFVVSFLIVAAGGSVLA |
| | Probability: | |
| 2413, | 0.994 AA1: | |
| 2414 | 24 AA2: 25 | MNFIIRYVFKRLLLLAPVLLAVSA |
| | Probability: | |
| 62147, | 0.998 AA1: | |
| 62148 | 27 AA2: 28 | MNFILRRIVIALITVFLVSIFCFLAFS |
| | Probability: | |
| 46845, | 0.975 AA1: | |
| 46846 | 27 AA2: 28 | MNFKERREKKAFLALANGSVFHGYSVG |
| | Probability: | |
| 27885, | 0.997 AA1: | |
| 27886 | 23 AA2: 24 | MNFKRIGLLLMTVLFLFGCPGEG |
| | Probability: | |
| 36857, | 0.703 AA1: | |
| 36858 | 28 AA2: 29 | MNFKRRFIVIFSLLAIFLLYLIIHYAVL |
| | Probability: | |
| 41375, | 0.995 AA1: | |
| 41376 | 18 AA2: 19 | MNFRIFFLLLLLPMAAIA |
| | Probability: | |
| 43177, | 0.993 AA1: | |
| 43178 | 15 AA2: 16 | MNFSIIFAVILLVLA |
| | Probability: | |
| 78159, | 0.966 AA1: | |
| 78160 | 23 AA2: 24 | MNFSTLLPRIATALVLLTVLGLA |
| 8753, | Probability: | |
| 8754 | 0.741 AA1: | MNFWEIFLIGVGLSMDAFA |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 50247, | 0.963 AA1: | |
| 50248 | 26 AA2: 27 | MNGEQIVKNFSLITIAAALCALLLFS |
| | Probability: | |
| 67961, | 0.995 AA1: | |
| 67962 | 24 AA2: 25 | MNGFCLAKKIAFLFLFFFVGISVA |
| | Probability: | |
| 19187, | 0.978 AA1: | |
| 19188 | 18 AA2: 19 | MNGFVLSLLMAAPQGAEG |
| | Probability: | |
| 50123, | 0.993 AA1: | |
| 50124 | 25 AA2: 26 | MNGIGKFSLTAGIGGLLISCMAGTA |
| | Probability: | |
| 22653, | 0.998 AA1: | |
| 22654 | 26 AA2: 27 | MNGIKRIITNTMFVTGAALVILSVFS |
| | Probability: | |
| 77481, | 1.000 AA1: | |
| 77482 | 25 AA2: 26 | MNGKTWIALLLTLSFTLFGVAAAIA |
| | Probability: | |
| 10115, | 0.596 AA1: | |
| 10116 | 17 AA2: 18 | MNGLAILIISAVVLSLA |
| | Probability: | |
| 23973, | 0.603 AA1: | |
| 23974 | 17 AA2: 18 | MNGLAILIISAVVLSLA |
| | Probability: | |
| 20755, | 0.963 AA1: | |
| 20756 | 24 AA2: 25 | MNGLRVSFLTAIFSSAARSTSTSK |
| | Probability: | |
| 53575, | 0.968 AA1: | |
| 53576 | 25 AA2: 26 | MNGMIYRRGKKRLIALAVLFWTLLA |
| | Probability: | |
| 52523, | 0.832 AA1: | |
| 52524 | 21 AA2: 22 | MNGNLKIGLLLTLLVVLISLT |

| | | |
|---|---|---|
| | Probability: | |
| 64049, | 0.954 AA1: | |
| 64050 | 22 AA2: 23 | MNGNNRKKIIPTIFLIAGIATA |
| | Probability: | |
| 69671, | 0.979 AA1: | |
| 69672 | 18 AA2: 19 | MNGRASFLCISPAPSLYA |
| | Probability: | |
| 51293, | 0.984 AA1: | |
| 51294 | 25 AA2: 26 | MNGSNYDVVVSPAGTVAAVPAASSA |
| | Probability: | |
| 22187, | 0.742 AA1: | |
| 22188 | 24 AA2: 25 | MNIFFVMSIICLAVCFIMFFYLKG |
| | Probability: | |
| 21615, | 0.814 AA1: | |
| 21616 | 16 AA2: 17 | MNIFILDACALLAVLA |
| | Probability: | |
| 42755, | 0.999 AA1: | |
| 42756 | 26 AA2: 27 | MNIFKQLCRTLVLSAVAVALSVGLMG |
| | Probability: | |
| 74139, | 0.982 AA1: | |
| 74140 | 25 AA2: 26 | MNIFLLWSHPIIQVIAALLGVYAMA |
| | Probability: | |
| 14409, | 0.781 AA1: | |
| 14410 | 24 AA2: 25 | MNIFLNTCSVCALTSPSTNFPVLE |
| | Probability: | |
| 4947, | 0.745 AA1: | |
| 4948 | 16 AA2: 17 | MNIGIIGLGLIGGSFA |
| | Probability: | |
| 51977, | 0.696 AA1: | |
| 51978 | 16 AA2: 17 | MNIGIIGLGLIGGSFA |
| | Probability: | |
| 64247, | 0.862 AA1: | |
| 64248 | 28 AA2: 29 | MNIGIIGLGLIGGSFAKAYKAAGHPVFA |
| 11187, | Probability: | MNIGIIGLGLIGGSFAKAYKAAGHTVYA |

| | | |
|---|---|---|
| 11188 | 0.784 AA1: 28 AA2: 29 Probability: | |
| 65679, 65680 | 0.783 AA1: 28 AA2: 29 Probability: | MNIGITGLGLIGGSFAKAYKAAGHTVFA |
| 20839, 20840 | 0.997 AA1: 26 AA2: 27 Probability: | MNIGTKTALFLVIIVLMLVITPTVWG |
| 28585, 28586 | 0.999 AA1: 22 AA2: 23 Probability: | MNIKTKLSIMMIAIVLAVTGGA |
| 75097, 75098 | 0.705 AA1: 17 AA2: 18 Probability: | MNILFALILIASGVVFL |
| 65743, 65744 | 0.819 AA1: 29 AA2: 30 Probability: | MNIMQKLSSKNLLVFNLVLIGAIFGFSLA |
| 40783, 40784 | 0.697 AA1: 23 AA2: 24 Probability: | MNINNLKIGAILGFIIVTMSLLS |
| 71389, 71390 | 0.501 AA1: 19 AA2: 20 Probability: | MNIPWKKNAALFISGQALS |
| 24243, 24244 | 0.997 AA1: 26 AA2: 27 Probability: | MNISINRNFVKFNYLVLAVIISFAAA |
| 41559, 41560 | 0.999 AA1: 22 AA2: 23 Probability: | MNITMAVKMMALLVLLAGVAAA |
| 25605, 25606 | 0.998 AA1: 16 AA2: 17 Probability: | MNKAATLLFLCTALFA |
| 22825, 22826 | 1.000 AA1: | MNKAVLMGMLFGIAASAVNA |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 67747, | 1.000 AA1: | |
| 67748 | 26 AA2: 27 | MNKCFKPFFVTAILAIFGIAGAQAPA |
| | Probability: | |
| 57585, | 0.999 AA1: | |
| 57586 | 26 AA2: 27 | MNKHIRRMAAVFVVAFASMISIVSYG |
| | Probability: | |
| 38857, | 0.888 AA1: | |
| 38858 | 21 AA2: 22 | MNKHKNGWALVLSGGGAMGLA |
| | Probability: | |
| 28739, | 0.999 AA1: | |
| 28740 | 16 AA2: 17 | MNKIIALLLLCTALFA |
| | Probability: | |
| 74645, | 0.955 AA1: | |
| 74646 | 21 AA2: 22 | MNKIIAVLSFCQFIMSAYCQV |
| | Probability: | |
| 45105, | 0.957 AA1: | |
| 45106 | 26 AA2: 27 | MNKIKKIINYIMFVTGTALVILTVFS |
| | Probability: | |
| 77175, | 0.998 AA1: | |
| 77176 | 21 AA2: 22 | MNKILLISLILIQTVFSQTQS |
| | Probability: | |
| 19801, | 1.000 AA1: | |
| 19802 | 20 AA2: 21 | MNKKALFVCVLLFASAALFA |
| | Probability: | |
| 15109, | 1.000 AA1: | |
| 15110 | 21 AA2: 22 | MNKKIIAVFVFFALVTVAANA |
| | Probability: | |
| 42603, | 0.666 AA1: | |
| 42604 | 27 AA2: 28 | MNKKKNKKIITSFLRFITVIIALILVL |
| | Probability: | |
| 1127, | 0.980 AA1: | |
| 1128 | 28 AA2: 29 | MNKKLISINKIICLXFVALCFLFITCNG |

| | | |
|---|---|---|
| | Probability: | |
| 473, | 0.994 AA1: | |
| 474 | 21 AA2: 22 | MNKKMVSLVFTALVCIGNIRA |
| | Probability: | |
| 61693, | 1.000 AA1: | |
| 61694 | 22 AA2: 23 | MNKKRSLFIFLLLFFISSALFA |
| | Probability: | |
| 43219, | 0.989 AA1: | |
| 43220 | 22 AA2: 23 | MNKKTVLCILILIMVYASCSQK |
| | Probability: | |
| 44635, | 1.000 AA1: | |
| 44636 | 26 AA2: 27 | MNKLNRLLFLFSLFLILALSSCSKLG |
| | Probability: | |
| 69077, | 1.000 AA1: | |
| 69078 | 26 AA2: 27 | MNKLNRLLFLFSLFLILALSSCSKLG |
| | Probability: | |
| 13051, | 0.999 AA1: | |
| 13052 | 24 AA2: 25 | MNKLYFALVLTLLFSACASTINIS |
| | Probability: | |
| 68173, | 0.999 AA1: | |
| 68174 | 22 AA2: 23 | MNKMIIAGLVFGVVTSMNSGWA |
| | Probability: | |
| 52073, | 0.987 AA1: | |
| 52074 | 23 AA2: 24 | MNKNFPFTFLRLFLVFFAGALMA |
| | Probability: | |
| 51685, | 1.000 AA1: | |
| 51686 | 21 AA2: 22 | MNKNIFIFLLFFLFFSLLLSA |
| | Probability: | |
| 71375, | 1.000 AA1: | |
| 71376 | 20 AA2: 21 | MNKNILCFIFMVFLGFSIFA |
| | Probability: | |
| 19337, | 0.999 AA1: | |
| 19338 | 24 AA2: 25 | MNKNLIQPIIVLALICLFVSGALS |
| 7859, | Probability: | MNKNLIQPIIVLSLICLFVSGALS |

| | | |
|---|---|---|
| 7860 | 0.998 AA1: | |
| | 24 AA2: 25 | |
| | Probability: | |
| 64313, | 1.000 AA1: | |
| 64314 | 20 AA2: 21 | MNKPIFCLILLLSLSQTARA |
| | Probability: | |
| 64007, | 1.000 AA1: | |
| 64008 | 21 AA2: 22 | MNKPLFFIFIILLSLSQTARA |
| | Probability: | |
| 24369, | 0.979 AA1: | |
| 24370 | 23 AA2: 24 | MNKQSLAKRVLQTAPLMILSVFA |
| | Probability: | |
| 29859, | 0.875 AA1: | |
| 29860 | 27 AA2: 28 | MNKRALRADILLLVTAAVWGFGFIAQR |
| | Probability: | |
| 51097, | 0.823 AA1: | |
| 51098 | 25 AA2: 26 | MNKRALRADVLLLLTAGIWGFGFVA |
| | Probability: | |
| 56445, | 1.000 AA1: | |
| 56446 | 25 AA2: 26 | MNKRISLALVLFFLVVTLCFGQAGA |
| | Probability: | |
| 43707, | 0.996 AA1: | |
| 43708 | 21 AA2: 22 | MNKSLLLLLALFVANCSNNNN |
| | Probability: | |
| 71101, | 0.923 AA1: | |
| 71102 | 27 AA2: 28 | MNKSLNMGTIIGVFGAIAMILASIILS |
| | Probability: | |
| 69967, | 1.000 AA1: | |
| 69968 | 20 AA2: 21 | MNKTIFCLIILLSLSQTARA |
| | Probability: | |
| 72069, | 1.000 AA1: | |
| 72070 | 25 AA2: 26 | MNKTKAIIVLLIAILIGVSTNNVHA |
| 35407, | Probability: | |
| 35408 | 0.540 AA1: | MNKTSLNFKIGETGAITASVTPANA |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 16719, | 0.603 AA1: | |
| 16720 | 26 AA2: 27 | MNKVISNVATQTNLLSMNAAIEAAHA |
| | Probability: | |
| 73855, | 0.511 AA1: | |
| 73856 | 26 AA2: 27 | MNKVMQIARKHNAIIVVAAGNENVLA |
| | Probability: | |
| 51947, | 0.999 AA1: | |
| 51948 | 18 AA2: 19 | MNKYFFVFLFFISFFVYA |
| | Probability: | |
| 7861, | 0.620 AA1: | |
| 7862 | 26 AA2: 27 | MNLAGLSFLFSLFALIISIFSFFYFR |
| | Probability: | |
| 62995, | 0.995 AA1: | |
| 62996 | 26 AA2: 27 | MNLFTRPGRGLLLAAVVIGAVCVSSA |
| | Probability: | |
| 71661, | 0.989 AA1: | |
| 71662 | 26 AA2: 27 | MNLGLIGAGLVMGISAVGSATGIAIA |
| | Probability: | |
| 3941, | 0.998 AA1: | |
| 3942 | 29 AA2: 30 | MNLGLLGAGLVLGIAAVGSGFGIGIAAQA |
| | Probability: | |
| 55437, | 0.979 AA1: | |
| 55438 | 28 AA2: 29 | MNLGLLGAGLVMGIAAIGSGIGIGYAGS |
| | Probability: | |
| 33197, | 0.952 AA1: | |
| 33198 | 26 AA2: 27 | MNLGLLGPGLVMGIAAVGSAIGIGFA |
| | Probability: | |
| 34445, | 0.858 AA1: | |
| 34446 | 29 AA2: 30 | MNLGTLIGVLGAMTMVIGSILLSGNSIMA |
| | Probability: | |
| 3457, | 0.522 AA1: | |
| 3458 | 20 AA2: 21 | MNLHGKKILLGVTGGIAAYK |

| | | |
|---|---|---|
| | Probability: | |
| 71247, | 0.952 AA1: | |
| 71248 | 24 AA2: 25 | MNLKKLLVFLIVFTACFSVFPQEK |
| | Probability: | |
| 34765, | 0.913 AA1: | |
| 34766 | 22 AA2: 23 | MNLKSIIPLAMLVASNVFMTFA |
| | Probability: | |
| 68285, | 0.824 AA1: | |
| 68286 | 28 AA2: 29 | MNLPKLSVTRRVTVTMLILIMVLFGAMS |
| | Probability: | |
| 45371, | 0.895 AA1: | |
| 45372 | 24 AA2: 25 | MNLRKIIIFFIFFAVCFSVFPQEK |
| | Probability: | |
| 8723, | 0.935 AA1: | |
| 8724 | 28 AA2: 29 | MNMGTIVGVFGAMAMIILSIILSGNVMS |
| | Probability: | |
| 45893, | 0.709 AA1: | |
| 45894 | 23 AA2: 24 | MNMILIAIISVTAIGGICGIILS |
| | Probability: | |
| 30615, | 1.000 AA1: | |
| 30616 | 25 AA2: 26 | MNMIRKCVFSVFLLFFACFSLFAQG |
| | Probability: | |
| 43223, | 0.985 AA1: | |
| 43224 | 21 AA2: 22 | MNMKKFVTFLIILLIIVFAFL |
| | Probability: | |
| 12509, | 1.000 AA1: | |
| 12510 | 26 AA2: 27 | MNMKKKRFFFFLFVILFFASLFNAWS |
| | Probability: | |
| 54173, | 0.999 AA1: | |
| 54174 | 25 AA2: 26 | MNMKKVLFFLAFLLFLVVSCSNENN |
| | Probability: | |
| 3179, | 0.633 AA1: | |
| 3180 | 22 AA2: 23 | MNMNRDLINRALLATGQTPLTA |
| 32227, | Probability: | MNMRLRLLLLLICSHCVFT |

| | | |
|---|---|---|
| 32228 | 0.994 AA1: | |
| | 19 AA2: 20 | |
| | Probability: | |
| 64411, | 0.998 AA1: | |
| 64412 | 22 AA2: 23 | MNNFVKFFACFLVAAVMAGCVS |
| | Probability: | |
| 42849, | 0.977 AA1: | |
| 42850 | 22 AA2: 23 | MNNFVKFFTCFLIAAVVAGCVS |
| | Probability: | |
| 67077, | 1.000 AA1: | |
| 67078 | 26 AA2: 27 | MNNFWKTHSRTTLIILAVVCAIGALA |
| | Probability: | |
| 28723, | 0.996 AA1: | |
| 28724 | 25 AA2: 26 | MNNIKIGVKLIASFLLVAMLSAAMG |
| | Probability: | |
| 72519, | 0.748 AA1: | |
| 72520 | 25 AA2: 26 | MNNKFLNGARRAGVMLALVLLGVSS |
| | Probability: | |
| 69427, | 0.965 AA1: | |
| 69428 | 21 AA2: 22 | MNNKIIIVGAGIAGLTAGIYA |
| | Probability: | |
| 63991, | 1.000 AA1: | |
| 63992 | 24 AA2: 25 | MNNKIRHLVFTALLLLIPCSFLQA |
| | Probability: | |
| 68857, | 0.910 AA1: | |
| 68858 | 22 AA2: 23 | MNNKKYTFLALIIGTGLAMIAA |
| | Probability: | |
| 50807, | 0.691 AA1: | |
| 50808 | 27 AA2: 28 | MNNKNFGIYIFLLVGVLFNCTTNTTQN |
| | Probability: | |
| 75039, | 0.596 AA1: | |
| 75040 | 18 AA2: 19 | MNNPCLILSMAIRFTSMA |
| 75123, | Probability: | |
| 75124 | 0.970 AA1: | MNNPTIKMCFGAFVVGGVIGALA |

| | | |
|---|---|---|
| | 23 AA2: 24 | |
| | Probability: | |
| 54975, | 1.000 AA1: | |
| 54976 | 25 AA2: 26 | MNNRTRIIIGAIVSLLIIVAAASLG |
| | Probability: | |
| 17841, | 0.997 AA1: | |
| 17842 | 21 AA2: 22 | MNNSNAWILFALLSAFFAALT |
| | Probability: | |
| 69729, | 0.997 AA1: | |
| 69730 | 26 AA2: 27 | MNNYTAAICVFSLLTIAGCTPSVRYA |
| | Probability: | |
| 76259, | 0.804 AA1: | |
| 76260 | 18 AA2: 19 | MNPGATLALSPFSEPLLS |
| | Probability: | |
| 20425, | 0.985 AA1: | |
| 20426 | 25 AA2: 26 | MNPGKKIIFVFFMMAVSVFYIPAQT |
| | Probability: | |
| 19721, | 1.000 AA1: | |
| 19722 | 19 AA2: 20 | MNPILSLIAFLLLPALLFA |
| | Probability: | |
| 60259, | 0.717 AA1: | |
| 60260 | 23 AA2: 24 | MNPLLLVPIGFGIVFANIPFANM |
| | Probability: | |
| 35117, | 0.958 AA1: | |
| 35118 | 21 AA2: 22 | MNPNLRFLLSGISAINTDFFA |
| | Probability: | |
| 40417, | 0.893 AA1: | |
| 40418 | 25 AA2: 26 | MNQFLQTAKILALVCFVSALLLSLM |
| | Probability: | |
| 31071, | 0.685 AA1: | |
| 31072 | 25 AA2: 26 | MNQRALRADILLLLTAGIWGFGFVA |
| | Probability: | |
| 40429, | 1.000 AA1: | |
| 40430 | 19 AA2: 20 | MNRACLLILALCCLCFALG |

| | | |
|---|---|---|
| | Probability: | |
| 64307, | 0.998 AA1: | |
| 64308 | 19 AA2: 20 | MNRFALMLLNIAAALYLFA |
| | Probability: | |
| 76623, | 0.914 AA1: | |
| 76624 | 22 AA2: 23 | MNRFLLFFLLTLPLFFFCSCKE |
| | Probability: | |
| 65407, | 0.897 AA1: | |
| 65408 | 21 AA2: 22 | MNRITGFFPAPYSIAASPLFS |
| | Probability: | |
| 57251, | 0.888 AA1: | |
| 57252 | 21 AA2: 22 | MNRITGFSPAPYSIAASPLFS |
| | Probability: | |
| 24115, | 0.689 AA1: | |
| 24116 | 25 AA2: 26 | MNRITRLRWSILATMGVGSRGCAML |
| | Probability: | |
| 22375, | 0.888 AA1: | |
| 22376 | 19 AA2: 20 | MNRKIFTTIFLILITGCGA |
| | Probability: | |
| 28821, | 0.811 AA1: | |
| 28822 | 27 AA2: 28 | MNRKNAPDFRPLMLVLLVFCGVSLLPA |
| | Probability: | |
| 53879, | 0.992 AA1: | |
| 53880 | 27 AA2: 28 | MNRKNAPDYRPLMLVLLVFCGASLLPA |
| | Probability: | |
| 43007, | 0.991 AA1: | |
| 43008 | 27 AA2: 28 | MNRKNASDFRPLMLVMLVFCGISLLPA |
| | Probability: | |
| 66831, | 0.740 AA1: | |
| 66832 | 26 AA2: 27 | MNRKSVIMTKTRRIALIIKLSGISIA |
| | Probability: | |
| 4987, | 0.995 AA1: | |
| 4988 | 26 AA2: 27 | MNRLTIRRLRGRTLLMTAATVMAACA |
| 26573, | Probability: | MNRLYFALVLTLLFSACASTINIS |

| | | |
|---|---|---|
| 26574 | 0.999 AA1: 24 AA2: 25 Probability: | |
| 18711, 18712 | 1.000 AA1: 22 AA2: 23 Probability: | MNRRFYLVLSIAILIGIQPVFA |
| 44083, 44084 | 0.805 AA1: 27 AA2: 28 Probability: | MNRRLILYIVSVFVFAIGLFLIQQSLN |
| 2771, 2772 | 0.807 AA1: 25 AA2: 26 Probability: | MNRVLFFLIPPAVGAVIGFITNAIA |
| 8237, 8238 | 0.867 AA1: 29 AA2: 30 Probability: | MNRYVKMITTDTQNHLMSAALAAAQFVSV |
| 60157, 60158 | 0.541 AA1: 28 AA2: 29 Probability: | MNSFSKFFVKLCFSIFLITALLPRSNFL |
| 16563, 16564 | 0.928 AA1: 29 AA2: 30 Probability: | MNSIIKLFSLFSKKEKSRLILIAVAVLFA |
| 7309, 7310 | 0.807 AA1: 22 AA2: 23 Probability: | MNSLAALPDLTILTFFMASPLA |
| 32589, 32590 | 0.514 AA1: 21 AA2: 22 Probability: | MNSLIVSCMGSTLKILFATLA |
| 42015, 42016 | 1.000 AA1: 24 AA2: 25 Probability: | MNSLKRIICLFAFAAVLVSGRLAA |
| 66805, 66806 | 0.679 AA1: 19 AA2: 20 Probability: | MNSLPKRRYIFAVSALLLA |
| 65361, 65362 | 1.000 AA1: | MNSLSIRKIIITVAFLTLTTHA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 27317, | 0.585 AA1: | |
| 27318 | 23 AA2: 24 | MNSSKLLEILRITLPVFAMLALG |
| | Probability: | |
| 40649, | 1.000 AA1: | |
| 40650 | 25 AA2: 26 | MNSSSAVCAVSAALVSSFSTSALWA |
| | Probability: | |
| 55251, | 0.601 AA1: | |
| 55252 | 15 AA2: 16 | MNTAPLTVAIPIVVA |
| | Probability: | |
| 11449, | 0.718 AA1: | |
| 11450 | 21 AA2: 22 | MNTGDSAAASANGLLTTVAAG |
| | Probability: | |
| 39777, | 0.586 AA1: | |
| 39778 | 26 AA2: 27 | MNTGLYNMLIAFLTALTIVLGMRIMG |
| | Probability: | |
| 60025, | 0.739 AA1: | |
| 60026 | 26 AA2: 27 | MNTIIHWVRKLGASCFFIGYFPIASG |
| | Probability: | |
| 29465, | 0.911 AA1: | |
| 29466 | 26 AA2: 27 | MNTIIYWTRKLGASCFFIGYFPFASG |
| | Probability: | |
| 13917, | 0.981 AA1: | |
| 13918 | 25 AA2: 26 | MNTILTVAAALSMALAWVGCGKDSK |
| | Probability: | |
| 57541, | 0.958 AA1: | |
| 57542 | 27 AA2: 28 | MNTLKSRLTVVAAFVIIAALGVSVFLN |
| | Probability: | |
| 2223, | 0.970 AA1: | |
| 2224 | 16 AA2: 17 | MNTMMIGLILVGFTLA |
| | Probability: | |
| 59879, | 0.830 AA1: | |
| 59880 | 18 AA2: 19 | MNTQIAVLGAGAWGTAIA |

| | | |
|---|---|---|
| | Probability: | |
| 69615, | 0.759 AA1: | |
| 69616 | 18 AA2: 19 | MNTQISVLGAGAWGTAIA |
| | Probability: | |
| 18769, | 0.999 AA1: | |
| 18770 | 24 AA2: 25 | MNTRAKILLSAAVVAIVAVLTVWA |
| | Probability: | |
| 36663, | 0.738 AA1: | |
| 36664 | 20 AA2: 21 | MNTRKLILSLAVNWPAKVLS |
| | Probability: | |
| 36859, | 0.574 AA1: | |
| 36860 | 21 AA2: 22 | MNVENAVAASAAAWVAGVSLK |
| | Probability: | |
| 55717, | 0.647 AA1: | |
| 55718 | 28 AA2: 29 | MNVGSGSGLTDGLIFAGLAVAVSMGAGA |
| | Probability: | |
| 56087, | 0.747 AA1: | |
| 56088 | 22 AA2: 23 | MNVKLVLNSTSAGNVGAVAGFA |
| | Probability: | |
| 75759, | 0.978 AA1: | |
| 75760 | 29 AA2: 30 | MNVKTRSAVNRGFILLTVTFSAFPAAVPS |
| | Probability: | |
| 66033, | 0.896 AA1: | |
| 66034 | 18 AA2: 19 | MNWYFFALCAIIAVSTFN |
| | Probability: | |
| 61839, | 0.998 AA1: | |
| 61840 | 25 AA2: 26 | MNYIIKPAASLFIAAAITVAALSFV |
| | Probability: | |
| 66643, | 0.998 AA1: | |
| 66644 | 25 AA2: 26 | MNYIIKPAASLFIAAAITVAALSFV |
| | Probability: | |
| 22159, | 0.993 AA1: | |
| 22160 | 27 AA2: 28 | MNYLKKRKITDGFWYALMTAASVAVGA |
| 53757, | Probability: | MNYLKKRKITDSFWYALMTAASIAVGA |

| | | |
|---|---|---|
| 53758 | 0.993 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 60513, | 0.993 AA1: | |
| 60514 | 27 AA2: 28 | MNYLKKRKVTDSLWYVLMAVASVAVGA |
| | Probability: | |
| 21313, | 0.952 AA1: | |
| 21314 | 29 AA2: 30 | MNYSNLPVFIAGAFLLGLSVFCSGCYTLK |
| | Probability: | |
| 56925, | 0.996 AA1: | |
| 56926 | 21 AA2: 22 | MPAAPAAAVPAATAPTAPAVA |
| | Probability: | |
| 50913, | 0.757 AA1: | |
| 50914 | 22 AA2: 23 | MPAGRLWRIPLITSLLPSSISA |
| | Probability: | |
| 11733, | 0.518 AA1: | |
| 11734 | 19 AA2: 20 | MPAIACAAALTVVVMVLQG |
| | Probability: | |
| 2345, | 0.699 AA1: | |
| 2346 | 18 AA2: 19 | MPAILNLIWLLPILAVFA |
| | Probability: | |
| 4039, | 0.798 AA1: | |
| 4040 | 22 AA2: 23 | MPAPAVSVFSIFRVLLTLAVIA |
| | Probability: | |
| 72831, | 0.970 AA1: | |
| 72832 | 28 AA2: 29 | MPARGALSTQAAPPCVVFMPSDSSASAA |
| | Probability: | |
| 4925, | 0.963 AA1: | |
| 4926 | 27 AA2: 28 | MPARSKSRRTAAYGRGGGTTPASSASA |
| | Probability: | |
| 38409, | 0.977 AA1: | |
| 38410 | 21 AA2: 22 | MPATFWLITFAPLAASPALRA |
| 37471, | Probability: | |
| 37472 | 0.681 AA1: | MPAVMISRNAAIFGSGLLIWHVHA |

| | | |
|---|---|---|
| | 24 AA2: 25 | |
| | Probability: | |
| 16001, | 0.986 AA1: | |
| 16002 | 21 AA2: 22 | MPAVTGSPSVAGLPAAPQALA |
| | Probability: | |
| 51831, | 0.729 AA1: | |
| 51832 | 24 AA2: 25 | MPCFRRKPQFILCSATIGNPKELA |
| | Probability: | |
| 14551, | 0.981 AA1: | |
| 14552 | 22 AA2: 23 | MPCRVSSSRCLGVATFATVTFA |
| | Probability: | |
| 70191, | 0.956 AA1: | |
| 70192 | 25 AA2: 26 | MPDLALSGFAFAFIFADFASFTGFA |
| | Probability: | |
| 49933, | 0.901 AA1: | |
| 49934 | 25 AA2: 26 | MPERISFIPSTVSLLVFMNASVSLT |
| | Probability: | |
| 64257, | 0.999 AA1: | |
| 64258 | 19 AA2: 20 | MPFICFSFTFFLFCVSVQA |
| | Probability: | |
| 14849, | 1.000 AA1: | |
| 14850 | 20 AA2: 21 | MPFLPTLIFYLLTLATTALA |
| | Probability: | |
| 71761, | 0.918 AA1: | |
| 71762 | 26 AA2: 27 | MPFLRLAAIISALLLPFMFFMPIGFA |
| | Probability: | |
| 56167, | 0.949 AA1: | |
| 56168 | 25 AA2: 26 | MPFNISFIPSTVSLLVLINSSVSLT |
| | Probability: | |
| 30171, | 0.975 AA1: | |
| 30172 | 27 AA2: 28 | MPFTRRCMKIKYIFLALCLCAFVNDYG |
| | Probability: | |
| 7579, | 0.979 AA1: | |
| 7580 | 19 AA2: 20 | MPGASDSLTVLGAAVSAFA |

| | | |
|---|---|---|
| | Probability: | |
| 36131, | 0.736 AA1: | |
| 36132 | 19 AA2: 20 | MPGCAAPRVLPAPAGSSFA |
| | Probability: | |
| 19491, | 0.972 AA1: | |
| 19492 | 22 AA2: 23 | MPGGIIIAAVIICAXFTSFTGA |
| | Probability: | |
| 38507, | 0.993 AA1: | |
| 38508 | 23 AA2: 24 | MPGGSSGGSAASVAACQCFASLG |
| | Probability: | |
| 5899, | 0.754 AA1: | |
| 5900 | 21 AA2: 22 | MPGKKRLLILGAGVLQGPALR |
| | Probability: | |
| 68221, | 0.655 AA1: | |
| 68222 | 24 AA2: 25 | MPGTGITRTAVVLLFTMPIAISSA |
| | Probability: | |
| 37953, | 0.966 AA1: | |
| 37954 | 27 AA2: 28 | MPGTGITRTAVVLLFTMPIAISSAIIA |
| | Probability: | |
| 44309, | 0.685 AA1: | |
| 44310 | 21 AA2: 22 | MPHFQSIANLSFGSSPLLVLG |
| | Probability: | |
| 17703, | 0.895 AA1: | |
| 17704 | 20 AA2: 21 | MPHHASFTIVSLAVSSFVFT |
| | Probability: | |
| 56677, | 0.910 AA1: | |
| 56678 | 16 AA2: 17 | MPIAISSAIIAAIVAA |
| | Probability: | |
| 32699, | 0.855 AA1: | |
| 32700 | 28 AA2: 29 | MPIRLRVILVIAFINLLIILFSVLFGTA |
| | Probability: | |
| 28475, | 0.553 AA1: | |
| 28476 | 23 AA2: 24 | MPISSWMRSSSISMCWRSFASRA |
| 1301, | Probability: | MPIVSALTTLSPGLATGVKSHS |

| | | |
|---|---|---|
| 1302 | 0.539 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 48011, | 0.782 AA1: | |
| 48012 | 26 AA2: 27 | MPKGVFMKRILVLTVCVVFLSVILAG |
| | Probability: | |
| 67973, | 0.630 AA1: | |
| 67974 | 20 AA2: 21 | MPKITFLGAGSTVFARNVLG |
| | Probability: | |
| 31543, | 0.999 AA1: | |
| 31544 | 17 AA2: 18 | MPKKFSTLLLLPALAFA |
| | Probability: | |
| 35433, | 0.851 AA1: | |
| 35434 | 22 AA2: 23 | MPKPAVWRTXVSALSISLAALG |
| | Probability: | |
| 21645, | 0.668 AA1: | |
| 21646 | 23 AA2: 24 | MPKRNDIKKILLIGSGPIVIGQA |
| | Probability: | |
| 53611, | 0.998 AA1: | |
| 53612 | 22 AA2: 23 | MPKSLLIIAAVALLFSCSPGGK |
| | Probability: | |
| 31969, | 0.774 AA1: | |
| 31970 | 18 AA2: 19 | MPLCAALPACPACGIAAG |
| | Probability: | |
| 32705, | 0.574 AA1: | |
| 32706 | 23 AA2: 24 | MPLDKNIKKIMVIGSGPIVIGQA |
| | Probability: | |
| 61907, | 0.577 AA1: | |
| 61908 | 23 AA2: 24 | MPLDKNIKKIMVIGSGPIVIGQA |
| | Probability: | |
| 55951, | 0.548 AA1: | |
| 55952 | 23 AA2: 24 | MPLDKNIKKTMVIGSGPIVIGQA |
| 33863, | Probability: | |
| 33864 | 0.571 AA1: | MPLDKNIKKVLVIGSGPIVIGQA |

| | | |
|---|---|---|
| | 23 AA2: 24 | |
| | Probability: | |
| 47853, | 0.571 AA1: | |
| 47854 | 23 AA2: 24 | MPLDKNIKKVLVIGSGPIVIGQA |
| | Probability: | |
| 70533, | 0.573 AA1: | |
| 70534 | 23 AA2: 24 | MPLDKNIKKVLVIGSGPIVIGQA |
| | Probability: | |
| 56431, | 0.558 AA1: | |
| 56432 | 19 AA2: 20 | MPLLCLLVFFLISCSKSDA |
| | Probability: | |
| 50441, | 0.987 AA1: | |
| 50442 | 27 AA2: 28 | MPLMRIALRSASSAARTIRLAAAIAAA |
| | Probability: | |
| 30707, | 0.865 AA1: | |
| 30708 | 21 AA2: 22 | MPMDSFEIFFALFVLVLLAVA |
| | Probability: | |
| 123, | 0.609 AA1: | |
| 124 | 23 AA2: 24 | MPMGYLIAGPVGTGKSFMVNAFA |
| | Probability: | |
| 69123, | 0.939 AA1: | |
| 69124 | 18 AA2: 19 | MPMHIMAAAAAAKGSSGA |
| | Probability: | |
| 22837, | 0.631 AA1: | |
| 22838 | 18 AA2: 19 | MPMHIMAVAAAARGSCGT |
| | Probability: | |
| 48843, | 0.998 AA1: | |
| 48844 | 23 AA2: 24 | MPMIILIAALILASGFFSLFEQA |
| | Probability: | |
| 27889, | 0.876 AA1: | |
| 27890 | 27 AA2: 28 | MPMIRSAEGRRLFVMGAKVVMAGSSIA |
| | Probability: | |
| 11331, | 0.797 AA1: | |
| 11332 | 30 AA2: 31 | MPMLAAIPFSLLASITTLLLLFTRSVPTIS |

| | | |
|---|---|---|
| | Probability: | |
| 40535, | 0.976 AA1: | |
| 40536 | 27 AA2: 28 | MPMNMTFLRYTLPLAAMLCISLLSSCA |
| | Probability: | |
| 38715, | 0.847 AA1: | |
| 38716 | 21 AA2: 22 | MPNGLTLSVVGVISGTPTTFG |
| | Probability: | |
| 54357, | 0.633 AA1: | |
| 54358 | 26 AA2: 27 | MPPAVVIASLIFAASSLETFSLRGLG |
| | Probability: | |
| 77703, | 0.648 AA1: | |
| 77704 | 29 AA2: 30 | MPPFIIIAKDIYSLIFLAPVPFTFSDASA |
| | Probability: | |
| 50259, | 0.591 AA1: | |
| 50260 | 27 AA2: 28 | MPPKPLIIFLCVIAIFFISLLIPVKNN |
| | Probability: | |
| 57689, | 0.996 AA1: | |
| 57690 | 30 AA2: 31 | MPPRLSILLTATMIGTPASFAWLIASMVWG |
| | Probability: | |
| 26337, | 0.990 AA1: | |
| 26338 | 21 AA2: 22 | MPPRPLALYALILILSLFLTA |
| | Probability: | |
| 34181, | 0.748 AA1: | |
| 34182 | 28 AA2: 29 | MPPVISCVIFSRFLLSFITSLLPSSISS |
| | Probability: | |
| 61325, | 0.971 AA1: | |
| 61326 | 26 AA2: 27 | MPQGAFMKKYILCFSIFCLCLPVIFA |
| | Probability: | |
| 26763, | 0.990 AA1: | |
| 26764 | 24 AA2: 25 | MPQIYNRKCKNSICLILAIVLAFA |
| | Probability: | |
| 9239, | 0.997 AA1: | |
| 9240 | 30 AA2: 31 | MPQVLIFPLFVVALMAALFVFLPGGVPLIA |
| 3975, | Probability: | MPQVLVYPAVILVLMMLLFMFMPKGSALA |

| | | |
|---|---|---|
| 3976 | 0.998 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| 54933, | 0.924 AA1: | |
| 54934 | 20 AA2: 21 | MPRFFILLLLALVLVGCGQK |
| | Probability: | |
| 67401, | 0.950 AA1: | |
| 67402 | 19 AA2: 20 | MPRLFCSSSLMICMARIFG |
| | Probability: | |
| 33671, | 0.947 AA1: | |
| 33672 | 19 AA2: 20 | MPRLFCSSSLMICMARILG |
| | Probability: | |
| 66415, | 0.525 AA1: | |
| 66416 | 19 AA2: 20 | MPRLFRSSSLMICMARILG |
| | Probability: | |
| 71005, | 0.810 AA1: | |
| 71006 | 28 AA2: 29 | MPRRAFLFARARSPLPLEAGFSSSPARA |
| | Probability: | |
| 75107, | 0.999 AA1: | |
| 75108 | 28 AA2: 29 | MPRTVLLSCLAGLGMLLLAMNPMLQSFE |
| | Probability: | |
| 47009, | 0.512 AA1: | |
| 47010 | 17 AA2: 18 | MPSANLLCSLPSIFAAS |
| | Probability: | |
| 6167, | 0.665 AA1: | |
| 6168 | 28 AA2: 29 | MPSANLLCSLPSIFAASNPLPISNPRVA |
| | Probability: | |
| 42445, | 0.565 AA1: | |
| 42446 | 28 AA2: 29 | MPSANLLCSLPVIFAASKPLPISNPLVA |
| | Probability: | |
| 69477, | 0.860 AA1: | |
| 69478 | 20 AA2: 21 | MPSASCVWFTFTSMYAPSNA |
| 71937, | Probability: | |
| 71938 | 0.553 AA1: | MPSESLPAMFLXLLRNFSIYA |

| | | |
|---|---|---|
| | 21 AA2: 22 | |
| | Probability: | |
| 12889, | 0.970 AA1: | |
| 12890 | 19 AA2: 20 | MPSGAAVMIFIGGGFPVHA |
| | Probability: | |
| 23851, | 1.000 AA1: | |
| 23852 | 25 AA2: 26 | MPSMRLAMSSALCALWLASSLISFA |
| | Probability: | |
| 37977, | 1.000 AA1: | |
| 37978 | 27 AA2: 28 | MPSMRRAMSSALCALWLASSLISPATT |
| | Probability: | |
| 17777, | 0.951 AA1: | |
| 17778 | 26 AA2: 27 | MPSPTVFIFLLLAACLAGFIDGVAGG |
| | Probability: | |
| 65325, | 0.653 AA1: | |
| 65326 | 23 AA2: 24 | MPSSFSRVVSSTSGFSPSPPVSV |
| | Probability: | |
| 67489, | 0.547 AA1: | |
| 67490 | 23 AA2: 24 | MPSSFSRVVSSTSGFSPSPPVSV |
| | Probability: | |
| 14049, | 0.902 AA1: | |
| 14050 | 19 AA2: 20 | MPSSHAALVSSLCTSIAFV |
| | Probability: | |
| 67011, | 0.988 AA1: | |
| 67012 | 27 AA2: 28 | MPSTTSSVVSMLLASSMVITPSLPTFS |
| | Probability: | |
| 74039, | 0.934 AA1: | |
| 74040 | 19 AA2: 20 | MPSTVVTVMVALPGPAAYT |
| | Probability: | |
| 68203, | 0.590 AA1: | |
| 68204 | 25 AA2: 26 | MPSVYITVGGSIASTVAAFPPARLM |
| | Probability: | |
| 42825, | 0.532 AA1: | |
| 42826 | 19 AA2: 20 | MPSYIFICILFTFAPIILS |

| | | |
|---|---|---|
| | Probability: | |
| 38509, | 0.949 AA1: | |
| 38510 | 21 AA2: 22 | MPTIVVVGIVFLAVIALIIFA |
| | Probability: | |
| 53579, | 0.706 AA1: | |
| 53580 | 20 AA2: 21 | MPTPAPTPAQPSVAAQPSVA |
| | Probability: | |
| 47523, | 0.909 AA1: | |
| 47524 | 18 AA2: 19 | MPTPKLSLVLACTLLFVA |
| | Probability: | |
| 77769, | 0.918 AA1: | |
| 77770 | 20 AA2: 21 | MPTPLRCTFTLAVSPGVSNS |
| | Probability: | |
| 5975, | 0.870 AA1: | |
| 5976 | 19 AA2: 20 | MPVFFVILLVLFNVFPVYS |
| | Probability: | |
| 31339, | 0.959 AA1: | |
| 31340 | 19 AA2: 20 | MPVFFVILLVLLNISPVYS |
| | Probability: | |
| 23021, | 0.723 AA1: | |
| 23022 | 19 AA2: 20 | MPVGSSASSSRGSFASALA |
| | Probability: | |
| 63419, | 0.784 AA1: | |
| 63420 | 22 AA2: 23 | MPVIKDCTFGIVGLGLMGGAIA |
| | Probability: | |
| 24083, | 0.633 AA1: | |
| 24084 | 18 AA2: 19 | MPVLGFVAFVVFMCVSTA |
| | Probability: | |
| 38257, | 0.574 AA1: | |
| 38258 | 17 AA2: 18 | MPVLLAFLSLCISSCNS |
| | Probability: | |
| 37681, | 0.512 AA1: | |
| 37682 | 27 AA2: 28 | MPVTIIPTVYRRFLPVICLLLVLLSLA |
| 22969, | Probability: | MPVWRHCMFKKLIFIGVLFLAVESLFA |

| | | |
|---|---|---|
| 22970 | 0.906 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 56387, | 0.651 AA1: | |
| 56388 | 24 AA2: 25 | MPWGRKPFLALTLLSARGTKYVTA |
| | Probability: | |
| 55949, | 0.989 AA1: | |
| 55950 | 28 AA2: 29 | MQAKTISIASFISFLSLVCLLFSLSSYL |
| | Probability: | |
| 11663, | 0.976 AA1: | |
| 11664 | 23 AA2: 24 | MQCVAAVLFLAASCAAGSNRRGS |
| | Probability: | |
| 63433, | 0.991 AA1: | |
| 63434 | 17 AA2: 18 | MQDRKILLLLLCTTLFA |
| | Probability: | |
| 31105, | 0.945 AA1: | |
| 31106 | 26 AA2: 27 | MQENKLDVKKTLLLSFGFFASSIAWS |
| | Probability: | |
| 52763, | 1.000 AA1: | |
| 52764 | 26 AA2: 27 | MQGTVKKNLFFMLCFLFIFIPVVVQA |
| | Probability: | |
| 12743, | 1.000 AA1: | |
| 12744 | 21 AA2: 22 | MQKAAVLFFILFLGFSVMLFA |
| | Probability: | |
| 33493, | 0.922 AA1: | |
| 33494 | 20 AA2: 21 | MQKIFGITAPALFFDTMLHA |
| | Probability: | |
| 22635, | 1.000 AA1: | |
| 22636 | 24 AA2: 25 | MQKKLKSALIVASIAASATLMVVA |
| | Probability: | |
| 5201, | 0.998 AA1: | |
| 5202 | 26 AA2: 27 | MQKNLKSLLVAASLVAGMVITFVACG |
| 20667, | Probability: | |
| 20668 | 0.978 AA1: | MQLAVLLPSAVVTVTVAFPAAT |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 44931, | 0.896 AA1: | |
| 44932 | 25 AA2: 26 | MQLKRFLNRKVLRVLLVLAVVKVLA |
| | Probability: | |
| 2101, | 1.000 AA1: | |
| 2102 | 30 AA2: 31 | MQLKSSSSSVAATGSSSSVSATGSSSSVSA |
| | Probability: | |
| 74235, | 0.955 AA1: | |
| 74236 | 24 AA2: 25 | MQLKYKIPLILFSAYSVIVGALIA |
| | Probability: | |
| 15723, | 0.793 AA1: | |
| 15724 | 27 AA2: 28 | MQLNGLVVAAIVVLAIIALVIFFYISS |
| | Probability: | |
| 4817, | 1.000 AA1: | |
| 4818 | 18 AA2: 19 | MQLTALLLAAAFAVVVAA |
| | Probability: | |
| 17767, | 0.967 AA1: | |
| 17768 | 29 AA2: 30 | MQMILFRQMAFKFASLLFCLLLFSCAVKT |
| | Probability: | |
| 70125, | 1.000 AA1: | |
| 70126 | 27 AA2: 28 | MQMKTFNPQKTLLVFICFLLVCTGLFA |
| | Probability: | |
| 54691, | 1.000 AA1: | |
| 54692 | 18 AA2: 19 | MQMKTILILFALAALVFA |
| | Probability: | |
| 54769, | 0.537 AA1: | |
| 54770 | 18 AA2: 19 | MQPTNLLSMNAAIEAAHA |
| | Probability: | |
| 7967, | 0.565 AA1: | |
| 7968 | 19 AA2: 20 | MQRIGIALALAQGCELLIA |
| | Probability: | |
| 24535, | 0.775 AA1: | |
| 24536 | 28 AA2: 29 | MQRKFIVFSSILFLLIFILGSVIFIFLM |

| | | |
|---|---|---|
| | Probability: | |
| 52675, | 0.948 AA1: | |
| 52676 | 21 AA2: 22 | MQRKGGVGKSLFAANLGVALA |
| | Probability: | |
| 74953, | 0.656 AA1: | |
| 74954 | 28 AA2: 29 | MQRRGCGTTTACASVEITVSSPSSPCGA |
| | Probability: | |
| 21731, | 0.569 AA1: | |
| 21732 | 29 AA2: 30 | MQSKTIPQIAGTLVIILALFMLIRTFTTS |
| | Probability: | |
| 607, | 0.868 AA1: | |
| 608 | 22 AA2: 23 | MQTLKNPLFSPATATRCGIAAT |
| | Probability: | |
| 57809, | 0.738 AA1: | |
| 57810 | 19 AA2: 20 | MQTLSTIAAAGCKTLYVSG |
| | Probability: | |
| 68087, | 0.927 AA1: | |
| 68088 | 19 AA2: 20 | MQTSCLPLIAISFTPPSLS |
| | Probability: | |
| 15955, | 1.000 AA1: | |
| 15956 | 26 AA2: 27 | MQVTVKKILFFVFCFLFIFLPVMAQA |
| | Probability: | |
| 67395, | 0.836 AA1: | |
| 67396 | 25 AA2: 26 | MQYIGMKRIRLIIILFTFSCIALHG |
| | Probability: | |
| 42979, | 0.973 AA1: | |
| 42980 | 25 AA2: 26 | MQYIGMKRLRLIIILFTFSCFALHG |
| | Probability: | |
| 29487, | 0.637 AA1: | |
| 29488 | 17 AA2: 18 | MRAFIVVGATTFGGGYA |
| | Probability: | |
| 4705, | 0.837 AA1: | |
| 4706 | 24 AA2: 25 | MRAGARSRSNCSSLIVLXMSPAEK |
| 18191, | Probability: | MRAGASFFSVSVLAGTSVDG |

| | | |
|---|---|---|
| 18192 | 0.575 AA1: | |
| | 20 AA2: 21 | |
| | Probability: | |
| 69915, | 1.000 AA1: | |
| 69916 | 20 AA2: 21 | MRALFMTAALIAVLVGTAAA |
| | Probability: | |
| 17445, | 0.574 AA1: | |
| 17446 | 20 AA2: 21 | MRALYLSPCLYTISMGMAVA |
| | Probability: | |
| 43399, | 0.823 AA1: | |
| 43400 | 26 AA2: 27 | MRARLHAFAAPNGTVFVCTGMLSRAA |
| | Probability: | |
| 42361, | 1.000 AA1: | |
| 42362 | 24 AA2: 25 | MRAVAKKKRSTVAAASAATAAVVA |
| | Probability: | |
| 76365, | 0.922 AA1: | |
| 76366 | 21 AA2: 22 | MRAWFLVTRALITSFTMPANN |
| | Probability: | |
| 20649, | 0.593 AA1: | |
| 20650 | 18 AA2: 19 | MRAYRAFLGGGLAVVDVA |
| | Probability: | |
| 75503, | 0.815 AA1: | |
| 75504 | 17 AA2: 18 | MRCVRAASIGIFSVAHG |
| | Probability: | |
| 56433, | 0.874 AA1: | |
| 56434 | 20 AA2: 21 | MRDKNGKVLTMAALRSQAQS |
| | Probability: | |
| 7811, | 0.923 AA1: | |
| 7812 | 27 AA2: 28 | MRDNFMLRLITLHSSLFTFHYSGSARA |
| | Probability: | |
| 77095, | 0.566 AA1: | |
| 77096 | 18 AA2: 19 | MRDVYPGLCIIAGNVATA |
| 25075, | Probability: | |
| 25076 | 0.722 AA1: | MREIRGVSRRKQLKGLVIAAGLGAIA |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 31205, | 0.949 AA1: | |
| 31206 | 27 AA2: 28 | MRERYVKKLLFVLILVTHLVSGAFCQS |
| | Probability: | |
| 24321, | 0.541 AA1: | |
| 24322 | 20 AA2: 21 | MRFAECFMLVPAFFLMLALR |
| | Probability: | |
| 13815, | 1.000 AA1: | |
| 13816 | 20 AA2: 21 | MRFFVSVAVVFLLMSLPATA |
| | Probability: | |
| 22211, | 1.000 AA1: | |
| 22212 | 20 AA2: 21 | MRFIFSAVICIFLSSFPVKA |
| | Probability: | |
| 72143, | 1.000 AA1: | |
| 72144 | 24 AA2: 25 | MRFIYKYLFLSALVLILIVPTLRA |
| | Probability: | |
| 16089, | 0.999 AA1: | |
| 16090 | 24 AA2: 25 | MRFKFSSLIWGTFLLLAAAFIVFN |
| | Probability: | |
| 21925, | 0.980 AA1: | |
| 21926 | 22 AA2: 23 | MRFLNISIIILLLLTACTKREA |
| | Probability: | |
| 30453, | 0.999 AA1: | |
| 30454 | 28 AA2: 29 | MRFPVTVISFTGLLLACPDSSSSSPVQV |
| | Probability: | |
| 25677, | 0.946 AA1: | |
| 25678 | 23 AA2: 24 | MRFRVIKYVLLILALAFTCACKN |
| | Probability: | |
| 70407, | 0.999 AA1: | |
| 70408 | 26 AA2: 27 | MRGAARRYAVSLFLLVCGLSPPSIVA |
| | Probability: | |
| 7199, | 1.000 AA1: | |
| 7200 | 25 AA2: 26 | MRGATRKYAILLSVLICGFSPSRSA |

| | | |
|---|---|---|
| | Probability: | |
| 37807, | 0.774 AA1: | |
| 37808 | 20 AA2: 21 | MRGFVSELMSMAALVWGLLA |
| | Probability: | |
| 20543, | 1.000 AA1: | |
| 20544 | 25 AA2: 26 | MRGKYIRIMLFCCSAMLCLTSTVYA |
| | Probability: | |
| 66479, | 0.973 AA1: | |
| 66480 | 15 AA2: 16 | MRGLLGLAFSPPSTA |
| | Probability: | |
| 61119, | 0.703 AA1: | |
| 61120 | 18 AA2: 19 | MRGLTWLIGSLSALGVVA |
| | Probability: | |
| 18093, | 0.683 AA1: | |
| 18094 | 28 AA2: 29 | MRGNQVVTGLVLTIFGTGVSGFIGKSLA |
| | Probability: | |
| 22029, | 1.000 AA1: | |
| 22030 | 26 AA2: 27 | MRGNRLKRFVFLSMLMLSGLFCACTA |
| | Probability: | |
| 47177, | 0.532 AA1: | |
| 47178 | 20 AA2: 21 | MRGQRAGLKIVFSSSAAVYG |
| | Probability: | |
| 40249, | 0.995 AA1: | |
| 40250 | 21 AA2: 22 | MRGSGIASFILPIMCSLPCLA |
| | Probability: | |
| 52319, | 0.535 AA1: | |
| 52320 | 22 AA2: 23 | MRGTLKALIGLALRILPRSTRS |
| | Probability: | |
| 67929, | 1.000 AA1: | |
| 67930 | 28 AA2: 29 | MRGTRIVNRKLTVLALFASFFILLPVSA |
| | Probability: | |
| 21687, | 0.755 AA1: | |
| 21688 | 25 AA2: 26 | MRGTSPREVLEAFIGALLPIPSVQA |
| 21651, | Probability: | MRHSIKAIKNSRITQALLVTGTSAAGA |

| | | |
|---|---|---|
| 21652 | 0.998 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 76357, | 0.661 AA1: | |
| 76358 | 23 AA2: 24 | MRHTVSASGIIFSGANPSSTTLS |
| | Probability: | |
| 35819, | 0.953 AA1: | |
| 35820 | 26 AA2: 27 | MRIESKSSQLLFTPIALTAFVMLASA |
| | Probability: | |
| 59507, | 0.733 AA1: | |
| 59508 | 25 AA2: 26 | MRIFCSLPVPRSFALTWTIPFASMS |
| | Probability: | |
| 37979, | 1.000 AA1: | |
| 37980 | 16 AA2: 17 | MRIFLAAIFVVSIAFA |
| | Probability: | |
| 62063, | 0.653 AA1: | |
| 62064 | 14 AA2: 15 | MRIFPGTAICVVLA |
| | Probability: | |
| 60693, | 0.618 AA1: | |
| 60694 | 17 AA2: 18 | MRIFPVLLIFLASSIFS |
| | Probability: | |
| 30997, | 0.992 AA1: | |
| 30998 | 16 AA2: 17 | MRIGFLILFASTLALA |
| | Probability: | |
| 2251, | 1.000 AA1: | |
| 2252 | 24 AA2: 25 | MRIIFLTVAFMATVWACIGGTVLA |
| | Probability: | |
| 44023, | 0.560 AA1: | |
| 44024 | 27 AA2: 28 | MRIIPIASGKGGTGKSLVAANLGVAFA |
| | Probability: | |
| 54549, | 0.590 AA1: | |
| 54550 | 27 AA2: 28 | MRIIPIASGKGGVGKSLVAANLGVAFA |
| 43549, | Probability: | |
| 43550 | 0.922 AA1: | MRIISPSAYLSTLIGLIRNFWA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 60525, | 1.000 AA1: | |
| 60526 | 24 AA2: 25 | MRIKFFASRFFFCLFLCFLITAAA |
| | Probability: | |
| 42009, | 0.999 AA1: | |
| 42010 | 20 AA2: 21 | MRIKFKLSIVVIAIMAVVVA |
| | Probability: | |
| 28021, | 0.995 AA1: | |
| 28022 | 26 AA2: 27 | MRIKFSYIVLAAALLMPAIVSCGFFG |
| | Probability: | |
| 49979, | 0.541 AA1: | |
| 49980 | 23 AA2: 24 | MRIKLAAAALPEAPPLIPAAFCA |
| | Probability: | |
| 13269, | 0.996 AA1: | |
| 13270 | 27 AA2: 28 | MRIKLKSAIKIICLFLLVFCCFFTACK |
| | Probability: | |
| 40217, | 0.999 AA1: | |
| 40218 | 24 AA2: 25 | MRILFSKIALTAFAVNSALSMAFA |
| | Probability: | |
| 2505, | 0.648 AA1: | |
| 2506 | 24 AA2: 25 | MRILIKYLKPFTLLLFLAVMILFL |
| | Probability: | |
| 62647, | 0.807 AA1: | |
| 62648 | 24 AA2: 25 | MRILKYVTKKRLIIVGALLVLLLA |
| | Probability: | |
| 56791, | 0.995 AA1: | |
| 56792 | 20 AA2: 21 | MRILYVCLFFIVFAILPTHA |
| | Probability: | |
| 48365, | 0.957 AA1: | |
| 48366 | 23 AA2: 24 | MRIPGRHQALNTAAGLALASSLA |
| | Probability: | |
| 14713, | 0.994 AA1: | |
| 14714 | 19 AA2: 20 | MRIRLAAPALIPAAFCAEA |

| | | |
|---|---|---|
| | Probability: | |
| 28559, | 0.955 AA1: | |
| 28560 | 27 AA2: 28 | MRISSIVFPFITFSFIMAVAVFWAVYA |
| | Probability: | |
| 5071, | 0.996 AA1: | |
| 5072 | 22 AA2: 23 | MRIVLVGFGLLAGSVASAIRQA |
| | Probability: | |
| 2043, | 1.000 AA1: | |
| 2044 | 20 AA2: 21 | MRKAGAFFLILLFGFTLLYA |
| | Probability: | |
| 63539, | 1.000 AA1: | |
| 63540 | 19 AA2: 20 | MRKAIIVLFALFVCTGLYA |
| | Probability: | |
| 69823, | 0.836 AA1: | |
| 69824 | 19 AA2: 20 | MRKCGLFLFGALFLLISCN |
| | Probability: | |
| 12967, | 0.996 AA1: | |
| 12968 | 23 AA2: 24 | MRKFALRITAVIIYICVIFSVHA |
| | Probability: | |
| 52391, | 1.000 AA1: | |
| 52392 | 25 AA2: 26 | MRKFLIFLLFPALLAVSCPSSPVSK |
| | Probability: | |
| 2245, | 0.967 AA1: | |
| 2246 | 27 AA2: 28 | MRKGDIILIIAVTLFAFPILYIAMLFA |
| | Probability: | |
| 9117, | 0.603 AA1: | |
| 9118 | 16 AA2: 17 | MRKGFACLFSGAPGTG |
| | Probability: | |
| 11029, | 0.542 AA1: | |
| 11030 | 23 AA2: 24 | MRKGISSVTIPDGVTSIGAKAFA |
| | Probability: | |
| 70913, | 0.999 AA1: | |
| 70914 | 25 AA2: 26 | MRKHSVTKRIIPVLICCFLSACALS |
| 51391, | Probability: | MRKIAVFTAFFLFAYALYA |

| | | |
|---|---|---|
| 51392 | 1.000 AA1: | |
| | 19 AA2: 20 | |
| | Probability: | |
| 30393, | 1.000 AA1: | |
| 30394 | 21 AA2: 22 | MRKIITKAIALLIAFSTVTLA |
| | Probability: | |
| 16983, | 0.924 AA1: | |
| 16984 | 19 AA2: 20 | MRKILKIVSIVFFLFFVMS |
| | Probability: | |
| 35745, | 0.921 AA1: | |
| 35746 | 25 AA2: 26 | MRKISFYFLLLILCGAHLFPQSILA |
| | Probability: | |
| 56211, | 0.996 AA1: | |
| 56212 | 24 AA2: 25 | MRKKALIKSLFFGIPALTAFITFA |
| | Probability: | |
| 53085, | 1.000 AA1: | |
| 53086 | 21 AA2: 22 | MRKKFSYIVLAAALIMPAVVS |
| | Probability: | |
| 42519, | 1.000 AA1: | |
| 42520 | 22 AA2: 23 | MRKKNLCGITAIAAVIVLTALA |
| | Probability: | |
| 69273, | 0.999 AA1: | |
| 69274 | 24 AA2: 25 | MRKKSFICLLSIFVISFLSVITNG |
| | Probability: | |
| 71287, | 1.000 AA1: | |
| 71288 | 24 AA2: 25 | MRKKSTLSLVGAAVALVCASAAVA |
| | Probability: | |
| 49047, | 0.943 AA1: | |
| 49048 | 18 AA2: 19 | MRKLFTAAPGTYSHSLMG |
| | Probability: | |
| 70851, | 1.000 AA1: | |
| 70852 | 20 AA2: 21 | MRKLILTAAAVAAICFAANA |
| 44181, | Probability: | |
| 44182 | 1.000 AA1: | MRKLILTTAAVAAAIAGAFA |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 70853, | 1.000 AA1: | |
| 70854 | 20 AA2: 21 | MRKLILTTAAVAAAIAGAFA |
| | Probability: | |
| 35339, | 0.994 AA1: | |
| 35340 | 21 AA2: 22 | MRKLILTTAAVAAAIGLAGCS |
| | Probability: | |
| 56443, | 0.994 AA1: | |
| 56444 | 21 AA2: 22 | MRKLILTTAAVAAAIGLAGCS |
| | Probability: | |
| 68779, | 0.994 AA1: | |
| 68780 | 21 AA2: 22 | MRKLILTTAAVAAAIGLAGCS |
| | Probability: | |
| 68781, | 0.994 AA1: | |
| 68782 | 21 AA2: 22 | MRKLILTTAAVAAAIGLAGCS |
| | Probability: | |
| 11389, | 0.987 AA1: | |
| 11390 | 25 AA2: 26 | MRKLILTTAAVAAAIGLAGCSKKPA |
| | Probability: | |
| 58777, | 0.999 AA1: | |
| 58778 | 27 AA2: 28 | MRKLILTTAAVAAAIGLAGCSMPPAPK |
| | Probability: | |
| 61945, | 0.985 AA1: | |
| 61946 | 22 AA2: 23 | MRKLILTTAAVAAAIWLAGCSK |
| | Probability: | |
| 27689, | 1.000 AA1: | |
| 27690 | 21 AA2: 22 | MRKLILTTAAVAALLVGTAAA |
| | Probability: | |
| 61947, | 1.000 AA1: | |
| 61948 | 20 AA2: 21 | MRKLILTTAAVTAAIAGAFA |
| | Probability: | |
| 56441, | 0.999 AA1: | |
| 56442 | 21 AA2: 22 | MRKLILTTAVVAAAIGLAAVV |

| | | |
|---|---|---|
| | Probability: | |
| 67921, | 0.968 AA1: | |
| 67922 | 22 AA2: 23 | MRKLILTTAVVAAAIWLAGCSK |
| | Probability: | |
| 32695, | 0.937 AA1: | |
| 32696 | 19 AA2: 20 | MRKLIVCVSLIFVLLFAVG |
| | Probability: | |
| 9103, | 1.000 AA1: | |
| 9104 | 22 AA2: 23 | MRKLLFLFILIFFSGGFLFSQA |
| | Probability: | |
| 27925, | 0.999 AA1: | |
| 27926 | 22 AA2: 23 | MRKLLFLLFPALAILAIFSCSS |
| | Probability: | |
| 25625, | 0.945 AA1: | |
| 25626 | 22 AA2: 23 | MRKLNQPLTLYLIIATVPTVIA |
| | Probability: | |
| 72539, | 0.995 AA1: | |
| 72540 | 24 AA2: 25 | MRKLSVKSVCGALLAVAAVGIAVV |
| | Probability: | |
| 35083, | 0.508 AA1: | |
| 35084 | 24 AA2: 25 | MRKLVKYLKPFTLLFILAVLILFL |
| | Probability: | |
| 35267, | 0.960 AA1: | |
| 35268 | 29 AA2: 30 | MRKLYTYPWLIVAITALITVFFAFQLPHA |
| | Probability: | |
| 72465, | 0.998 AA1: | |
| 72466 | 22 AA2: 23 | MRKNVKAFLIIAAVMLASNCES |
| | Probability: | |
| 71233, | 0.987 AA1: | |
| 71234 | 22 AA2: 23 | MRKNVKAFLVIAVVLLAANCDK |
| | Probability: | |
| 66625, | 1.000 AA1: | |
| 66626 | 25 AA2: 26 | MRKQFSLVFFILLAVCFLGALPFYX |
| 57965, | Probability: | MRKQFSLVFLILLAVCILGALPFYFSG |

| | | |
|---|---|---|
| 57966 | 0.997 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 1549, | 0.998 AA1: | |
| 1550 | 25 AA2: 26 | MRKQLVTKRIIPVLICCFLSACALS |
| | Probability: | |
| 54783, | 1.000 AA1: | |
| 54784 | 23 AA2: 24 | MRKRHMALLLTVLFFILTGLLSA |
| | Probability: | |
| 31731, | 1.000 AA1: | |
| 31732 | 24 AA2: 25 | MRKTEKVFLFLLVAITIAICPGFA |
| | Probability: | |
| 69625, | 1.000 AA1: | |
| 69626 | 19 AA2: 20 | MRKTFIFSAFFLVISALYA |
| | Probability: | |
| 74275, | 0.999 AA1: | |
| 74276 | 23 AA2: 24 | MRKTFPRCSALSANFSSFTTSKA |
| | Probability: | |
| 54079, | 1.000 AA1: | |
| 54080 | 27 AA2: 28 | MRKTKIDASAFLLILIVLLLVGGVVFA |
| | Probability: | |
| 54437, | 1.000 AA1: | |
| 54438 | 26 AA2: 27 | MRKTTIIILLLAAVFLSLFSCNGSFL |
| | Probability: | |
| 28971, | 1.000 AA1: | |
| 28972 | 20 AA2: 21 | MRKVFLILILVLITAGFAAA |
| | Probability: | |
| 40415, | 0.999 AA1: | |
| 40416 | 21 AA2: 22 | MRKVVLAAMLIVSVMLVVSCA |
| | Probability: | |
| 77129, | 1.000 AA1: | |
| 77130 | 23 AA2: 24 | MRKYLSTALLALSLSLASLSCSG |
| 20143, | Probability: | |
| 20144 | 0.859 AA1: | MRLEGKTALISGAAQGLGAAMARRFA |

| | | |
|---|---|---|
| | 26 AA2: 27 | |
| | Probability: | |
| 25783, | 0.673 AA1: | |
| 25784 | 22 AA2: 23 | MRLENKKALVTGALRGIGKAIA |
| | Probability: | |
| 78039, | 0.999 AA1: | |
| 78040 | 21 AA2: 22 | MRLFSSTCASSLSSLITSRAA |
| | Probability: | |
| 10141, | 0.883 AA1: | |
| 10142 | 26 AA2: 27 | MRLGMKTKFIFITGGVLSSLGKGLTA |
| | Probability: | |
| 9737, | 0.993 AA1: | |
| 9738 | 20 AA2: 21 | MRLKIFALVCLVLICGCSRS |
| | Probability: | |
| 40107, | 0.904 AA1: | |
| 40108 | 27 AA2: 28 | MRLKMKPTKKSRLFTAALLLALCTVFS |
| | Probability: | |
| 50843, | 0.985 AA1: | |
| 50844 | 21 AA2: 22 | MRLKQYCCLLLLFTVKTVCFA |
| | Probability: | |
| 65215, | 1.000 AA1: | |
| 65216 | 26 AA2: 27 | MRLKSSILSIICISFTFFLFCVSAQA |
| | Probability: | |
| 26051, | 0.723 AA1: | |
| 26052 | 16 AA2: 17 | MRLKSVITICGMLTLA |
| | Probability: | |
| 10499, | 0.678 AA1: | |
| 10500 | 24 AA2: 25 | MRLKYLGKKGLLTDLLKAMGSLSA |
| | Probability: | |
| 17307, | 0.544 AA1: | |
| 17308 | 22 AA2: 23 | MRLLPSRAAFNAPLATLCLFLA |
| | Probability: | |
| 71805, | 0.971 AA1: | |
| 71806 | 22 AA2: 23 | MRLLPSRAALSAPLATLCLFLA |

| | | |
|---|---|---|
| | Probability: | |
| 67119, | 0.690 AA1: | |
| 67120 | 24 AA2: 25 | MRLMNVVDVAVIGGGPAGLAAAIA |
| | Probability: | |
| 22511, | 0.797 AA1: | |
| 22512 | 29 AA2: 30 | MRLPKMDTLDGKRPALCALSRAHMGSTAA |
| | Probability: | |
| 74621, | 0.943 AA1: | |
| 74622 | 27 AA2: 28 | MRLPWGKSSLALPCLPGTAIFISLEAA |
| | Probability: | |
| 72495, | 0.944 AA1: | |
| 72496 | 20 AA2: 21 | MRLRQKLSILAFLFFQYLLS |
| | Probability: | |
| 67161, | 0.773 AA1: | |
| 67162 | 23 AA2: 24 | MRLRSIVTICGMLTLAIGGYGCG |
| | Probability: | |
| 47165, | 0.807 AA1: | |
| 47166 | 21 AA2: 22 | MRLSFFTQSFSSAALSTSTSK |
| | Probability: | |
| 23359, | 0.869 AA1: | |
| 23360 | 26 AA2: 27 | MRLSIISSFSFRACISIPFIVPQSSS |
| | Probability: | |
| 60429, | 0.990 AA1: | |
| 60430 | 24 AA2: 25 | MRLTPVFTIFLLLLFFACSGEKNT |
| | Probability: | |
| 14397, | 0.628 AA1: | |
| 14398 | 20 AA2: 21 | MRLVKVTALIXIFVSSLFHT |
| | Probability: | |
| 48575, | 0.801 AA1: | |
| 48576 | 26 AA2: 27 | MRMAKYLFSLWAGVLIYASLFLITGA |
| | Probability: | |
| 24127, | 0.615 AA1: | |
| 24128 | 24 AA2: 25 | MRMFSSCSATVFTFVSIDSNLVVT |
| 75543, | Probability: | MRMKKLLLFLCCAASFLCRA |

| | | |
|---|---|---|
| 75544 | 1.000 AA1: | |
| | 20 AA2: 21 | |
| | Probability: | |
| 27133, | 0.955 AA1: | |
| 27134 | 21 AA2: 22 | MRMLFHCCCGPCATACVESLS |
| | Probability: | |
| 53567, | 0.996 AA1: | |
| 53568 | 26 AA2: 27 | MRMNLWRIAIIILLALGLLLSTVRYS |
| | Probability: | |
| 23187, | 1.000 AA1: | |
| 23188 | 29 AA2: 30 | MRMQFSKLALAATFGLAMAFTFTACMTHL |
| | Probability: | |
| 68375, | 0.716 AA1: | |
| 68376 | 25 AA2: 26 | MRNAYSFKILFAFGIALIYFLVCRA |
| | Probability: | |
| 6047, | 1.000 AA1: | |
| 6048 | 20 AA2: 21 | MRNFFIFSFILCALASQAFA |
| | Probability: | |
| 5495, | 1.000 AA1: | |
| 5496 | 19 AA2: 20 | MRNIASIAILAACIAVVGA |
| | Probability: | |
| 53171, | 1.000 AA1: | |
| 53172 | 23 AA2: 24 | MRNKLTLAVASGLALAFTLSCSS |
| | Probability: | |
| 43087, | 0.922 AA1: | |
| 43088 | 18 AA2: 19 | MRNLTCIVCPIGCSLAVN |
| | Probability: | |
| 65969, | 0.708 AA1: | |
| 65970 | 19 AA2: 20 | MRNLTCIVCPIGCSLTVNE |
| | Probability: | |
| 61901, | 0.604 AA1: | |
| 61902 | 17 AA2: 18 | MRNLTCIVCPMGCSLEA |
| 76913, | Probability: | |
| 76914 | 0.997 AA1: | MRNLVAIILLFLPALA |

| | | |
|---|---|---|
| | 16 AA2: 17 | |
| | Probability: | |
| 69831, | 0.973 AA1: | |
| 69832 | 16 AA2: 17 | MRNMLLFVTVAGVTVA |
| | Probability: | |
| 25611, | 0.998 AA1: | |
| 25612 | 23 AA2: 24 | MRNNFFVMGMVTLALAFAFIFTG |
| | Probability: | |
| 11801, | 0.975 AA1: | |
| 11802 | 20 AA2: 21 | MRNRFASPYIXGVSSGASLG |
| | Probability: | |
| 57371, | 0.965 AA1: | |
| 57372 | 23 AA2: 24 | MRNRGLFLFWAFVLLFSCNPSGR |
| | Probability: | |
| 69633, | 0.998 AA1: | |
| 69634 | 23 AA2: 24 | MRNSRILSAFFSSRSSFSFSAFS |
| | Probability: | |
| 60757, | 0.809 AA1: | |
| 60758 | 21 AA2: 22 | MRNXQRALLFILLFGTMLIVG |
| | Probability: | |
| 19355, | 1.000 AA1: | |
| 19356 | 24 AA2: 25 | MRPFNSLVALPISVLLFANTMALA |
| | Probability: | |
| 57099, | 1.000 AA1: | |
| 57100 | 24 AA2: 25 | MRPFNSLVALPISVLLFANTMALA |
| | Probability: | |
| 76251, | 0.990 AA1: | |
| 76252 | 26 AA2: 27 | MRPFQLAPGGGLCGAIFAAAGRKALA |
| | Probability: | |
| 30037, | 0.999 AA1: | |
| 30038 | 21 AA2: 22 | MRPISLACLTVVLILTGACSA |
| | Probability: | |
| 63809, | 0.673 AA1: | |
| 63810 | 17 AA2: 18 | MRPKNILTQAAFMNALA |

| | | |
|---|---|---|
| | Probability: | |
| 77321, | 0.946 AA1: | |
| 77322 | 24 AA2: 25 | MRPLLVLQIMLVIFSLAGALIKVA |
| | Probability: | |
| 77401, | 0.955 AA1: | |
| 77402 | 25 AA2: 26 | MRPLMLAMMIGMLLFTLLKPQLGAA |
| | Probability: | |
| 16537, | 0.884 AA1: | |
| 16538 | 22 AA2: 23 | MRPLVNMAFARLVAADASTDAA |
| | Probability: | |
| 44015, | 0.657 AA1: | |
| 44016 | 13 AA2: 14 | MRPSAPTAIAAFA |
| | Probability: | |
| 28745, | 0.899 AA1: | |
| 28746 | 24 AA2: 25 | MRPSILIIHTGGTLGMAAASGSQG |
| | Probability: | |
| 26493, | 0.999 AA1: | |
| 26494 | 29 AA2: 30 | MRQEWKTGFALFSACAALILLAGAAYYRA |
| | Probability: | |
| 72265, | 0.969 AA1: | |
| 72266 | 27 AA2: 28 | MRQFKEIIIAILIVTAVSFLLYSLVQS |
| | Probability: | |
| 70285, | 0.908 AA1: | |
| 70286 | 23 AA2: 24 | MRQFKEIVIAIFVVAAVSFLLYS |
| | Probability: | |
| 61203, | 0.605 AA1: | |
| 61204 | 16 AA2: 17 | MRQKLAYIIFCLLLLS |
| | Probability: | |
| 41669, | 0.996 AA1: | |
| 41670 | 21 AA2: 22 | MRQKLCLLACIVLLCGCSRNG |
| | Probability: | |
| 70729, | 0.999 AA1: | |
| 70730 | 20 AA2: 21 | MRQKLCLLACLVLICGCSRS |
| 57039, | Probability: | MRQMSERRKALAAIIACVLFWGFSFIS |

| | | |
|---|---|---|
| 57040 | 0.732 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 63347, | 0.953 AA1: | |
| 63348 | 26 AA2: 27 | MRQRWRSVAIRASAGAAWLSSARAVR |
| | Probability: | |
| 53343, | 0.732 AA1: | |
| 53344 | 18 AA2: 19 | MRRAENLASLTTLAAGVA |
| | Probability: | |
| 76213, | 0.918 AA1: | |
| 76214 | 22 AA2: 23 | MRRCLTAFFIALLVITAITCKS |
| | Probability: | |
| 55983, | 0.871 AA1: | |
| 55984 | 22 AA2: 23 | MRRFFAFLTIFAPSTTMRSVSR |
| | Probability: | |
| 53679, | 1.000 AA1: | |
| 53680 | 23 AA2: 24 | MRRFGIIICLFFILSFFPAALSA |
| | Probability: | |
| 71657, | 1.000 AA1: | |
| 71658 | 23 AA2: 24 | MRRFGIIICLFFILSFFPAALSA |
| | Probability: | |
| 76499, | 0.945 AA1: | |
| 76500 | 21 AA2: 22 | MRRFSTAASCSASLTRSPRPS |
| | Probability: | |
| 36461, | 0.999 AA1: | |
| 36462 | 23 AA2: 24 | MRRKIPFIAASLLALAITFVTCG |
| | Probability: | |
| 76605, | 0.999 AA1: | |
| 76606 | 26 AA2: 27 | MRRLTCAFALAAIIAGLSGVWQFATG |
| | Probability: | |
| 67385, | 0.977 AA1: | |
| 67386 | 25 AA2: 26 | MRRNFMKNYFLCFVMILSLLTAVCG |
| 16501, | Probability: | |
| 16502 | 0.942 AA1: | MRRNLFLKLIVTISLTTLLANCGLG |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 46053, | 0.812 AA1: | |
| 46054 | 26 AA2: 27 | MRRNTTTAKLNLILYLAVVLVSLGIA |
| | Probability: | |
| 77951, | 0.720 AA1: | |
| 77952 | 28 AA2: 29 | MRRPRSPLKSSRPAVFSGLFLFCFQGAR |
| | Probability: | |
| 6025, | 0.643 AA1: | |
| 6026 | 22 AA2: 23 | MRRSSVSTRSTLAIVLPALLHV |
| | Probability: | |
| 5859, | 1.000 AA1: | |
| 5860 | 21 AA2: 22 | MRRSVLCVTALFVFFSFSAFS |
| | Probability: | |
| 7395, | 0.973 AA1: | |
| 7396 | 20 AA2: 21 | MRRVFLLLIIHFFSCGLVFS |
| | Probability: | |
| 30501, | 0.999 AA1: | |
| 30502 | 20 AA2: 21 | MRRVLLCLLIVWALAAVCCA |
| | Probability: | |
| 51533, | 0.627 AA1: | |
| 51534 | 16 AA2: 17 | MRRVVVTGGSIVSALG |
| | Probability: | |
| 35311, | 1.000 AA1: | |
| 35312 | 20 AA2: 21 | MRRYLLLAISFLFCGSFLFA |
| | Probability: | |
| 65463, | 0.734 AA1: | |
| 65464 | 25 AA2: 26 | MRRYSGKLPFSSWILIIIPPLVLWA |
| | Probability: | |
| 60211, | 1.000 AA1: | |
| 60212 | 20 AA2: 21 | MRSIRQILTAVAAAPMAALS |
| | Probability: | |
| 37357, | 0.957 AA1: | |
| 37358 | 21 AA2: 22 | MRSLLTMLGIIIGVASVVVMV |

| | | |
|---|---|---|
| | Probability: | |
| 53321, | 1.000 AA1: | |
| 53322 | 30 AA2: 31 | MRSLRTKILTIVLAFLMFVSAAFVIYSLFA |
| | Probability: | |
| 24051, | 0.997 AA1: | |
| 24052 | 21 AA2: 22 | MRSLRTKILTIVLTFLMLVSA |
| | Probability: | |
| 21137, | 1.000 AA1: | |
| 21138 | 26 AA2: 27 | MRSNISTLALTAALMLAITFTLSCSG |
| | Probability: | |
| 26975, | 0.918 AA1: | |
| 26976 | 26 AA2: 27 | MRSPLSEAAFIAFLMSSYLAGFSRSA |
| | Probability: | |
| 19611, | 0.807 AA1: | |
| 19612 | 28 AA2: 29 | MRSQVSPVWAGFVXFFIWLCGYATPLFS |
| | Probability: | |
| 66023, | 0.894 AA1: | |
| 66024 | 24 AA2: 25 | MRSRSSASRSIVSISLSIPIFCKA |
| | Probability: | |
| 2061, | 0.918 AA1: | |
| 2062 | 20 AA2: 21 | MRSRTHILIALGAASLMLLS |
| | Probability: | |
| 19017, | 0.839 AA1: | |
| 19018 | 19 AA2: 20 | MRSSAVAALGPFSGSNVEN |
| | Probability: | |
| 71623, | 0.917 AA1: | |
| 71624 | 20 AA2: 21 | MRSVRNGTCTASILLFQAAA |
| | Probability: | |
| 5963, | 0.990 AA1: | |
| 5964 | 19 AA2: 20 | MRTDCVSAACTALPSGGLA |
| | Probability: | |
| 61965, | 0.985 AA1: | |
| 61966 | 22 AA2: 23 | MRTFIWCPCAVSTTSISTPAST |
| 42951, | Probability: | MRTGFFLPRSAGVPLIALLAFSIASTA |

| | | |
|---|---|---|
| 42952 | 1.000 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 72215, | 0.994 AA1: | |
| 72216 | 16 AA2: 17 | MRTILSCAAVSGVVKA |
| | Probability: | |
| 56423, | 1.000 AA1: | |
| 56424 | 23 AA2: 24 | MRTIRFFLLLTLIFSFMPVTAGA |
| | Probability: | |
| 15085, | 0.911 AA1: | |
| 15086 | 20 AA2: 21 | MRTKITPLFTILIFSFLIFS |
| | Probability: | |
| 65573, | 0.995 AA1: | |
| 65574 | 24 AA2: 25 | MRTLKSFFFLLFLPFLAVFISCSA |
| | Probability: | |
| 63607, | 1.000 AA1: | |
| 63608 | 26 AA2: 27 | MRTQFSKLALVAAFGLAMAFTFGCSS |
| | Probability: | |
| 61061, | 0.988 AA1: | |
| 61062 | 21 AA2: 22 | MRTRFAFSGIAPAPMPAAFCA |
| | Probability: | |
| 69683, | 0.997 AA1: | |
| 69684 | 27 AA2: 28 | MRTVHFAKIAGIVLLPVFLLSCFMAAG |
| | Probability: | |
| 61503, | 0.971 AA1: | |
| 61504 | 21 AA2: 22 | MRTVSSSSTTMTFFFSLPVST |
| | Probability: | |
| 29799, | 0.996 AA1: | |
| 29800 | 27 AA2: 28 | MRTYHSVPTVVVIALAGLIALSCASTG |
| | Probability: | |
| 77041, | 0.928 AA1: | |
| 77042 | 26 AA2: 27 | MRVFPSRMAARAISGAIFGSGGALYA |
| 10729, | Probability: | |
| 10730 | 0.513 AA1: | MRVGIVVCAGMSFVITPPRVS |

| | | |
|---|---|---|
| | 21 AA2: 22 | |
| | Probability: | |
| 72489, | 1.000 AA1: | |
| 72490 | 18 AA2: 19 | MRVIFPLLLIFFAVNVFA |
| | Probability: | |
| 44971, | 0.594 AA1: | |
| 44972 | 20 AA2: 21 | MRVLFAASPAIAIPALESLT |
| | Probability: | |
| 8501, | 0.757 AA1: | |
| 8502 | 29 AA2: 30 | MRVMTSMVFALLCFTGXFVMFFHIMRSLE |
| | Probability: | |
| 25459, | 0.561 AA1: | |
| 25460 | 22 AA2: 23 | MRVSARLSWRISLLPTSRSGLS |
| | Probability: | |
| 9333, | 0.641 AA1: | |
| 9334 | 21 AA2: 22 | MRVVLSVFVGLCLVFIVSLYG |
| | Probability: | |
| 1887, | 0.989 AA1: | |
| 1888 | 24 AA2: 25 | MRYFVRLAAPIMLVAFIAAIISNV |
| | Probability: | |
| 69327, | 0.995 AA1: | |
| 69328 | 25 AA2: 26 | MRYKFTPLIVFLALFLASCMSINHS |
| | Probability: | |
| 69719, | 0.699 AA1: | |
| 69720 | 18 AA2: 19 | MRYKTIYLLFMVFFLLXS |
| | Probability: | |
| 50257, | 0.818 AA1: | |
| 50258 | 24 AA2: 25 | MRYSTSRFLINSSSFFNSPSSSFS |
| | Probability: | |
| 27935, | 0.993 AA1: | |
| 27936 | 22 AA2: 23 | MSADIFSGILLSLAPGALLANA |
| | Probability: | |
| 23087, | 0.997 AA1: | |
| 23088 | 27 AA2: 28 | MSAFLKPLFALVISILLFAGLIFLANG |

| | | |
|---|---|---|
| | Probability: | |
| 65663, | 1.000 AA1: | |
| 65664 | 20 AA2: 21 | MSAITALTALIALTAVTATA |
| | Probability: | |
| 23243, | 0.845 AA1: | |
| 23244 | 17 AA2: 18 | MSALFAILLTVPAGVLS |
| | Probability: | |
| 59795, | 0.994 AA1: | |
| 59796 | 18 AA2: 19 | MSANVFSTGLFLPAQAAA |
| | Probability: | |
| 65861, | 0.994 AA1: | |
| 65862 | 18 AA2: 19 | MSANVFSTGLFLPAQAAA |
| | Probability: | |
| 70683, | 0.603 AA1: | |
| 70684 | 24 AA2: 25 | MSARATNFVLIGSFAPARFIASSA |
| | Probability: | |
| 56567, | 1.000 AA1: | |
| 56568 | 24 AA2: 25 | MSARLILPALLLLLLALLPQAAYA |
| | Probability: | |
| 17391, | 0.952 AA1: | |
| 17392 | 20 AA2: 21 | MSASSVWATTRIGMPQSAWA |
| | Probability: | |
| 29047, | 0.620 AA1: | |
| 29048 | 23 AA2: 24 | MSAVGAGPVGMLIAFVVSMIVCA |
| | Probability: | |
| 4533, | 0.967 AA1: | |
| 4534 | 25 AA2: 26 | MSCGASCGASSATFSCGASCGTSFA |
| | Probability: | |
| 76725, | 1.000 AA1: | |
| 76726 | 22 AA2: 23 | MSCMYPRLLLLAILAALPLCFA |
| | Probability: | |
| 5553, | 0.739 AA1: | |
| 5554 | 19 AA2: 20 | MSDKKRVVITAGSIISALG |
| 3687, | Probability: | MSDLATSSVILTPPSAFFTDSSIMFAVFFA |

| | | |
|---|---|---|
| 3688 | 0.902 AA1: | |
| | 30 AA2: 31 | |
| | Probability: | |
| 59195, | 0.570 AA1: | |
| 59196 | 22 AA2: 23 | MSEFIDRPRSSCALGGALAAIS |
| | Probability: | |
| 9627, | 0.725 AA1: | |
| 9628 | 24 AA2: 25 | MSEKKGGFAGMLKLGVVLALYAVA |
| | Probability: | |
| 61483, | 0.944 AA1: | |
| 61484 | 23 AA2: 24 | MSELKKLKLFIIILIFSANCLFS |
| | Probability: | |
| 51945, | 0.969 AA1: | |
| 51946 | 28 AA2: 29 | MSERTIMHKRTILPTLIGTALLLAGQSA |
| | Probability: | |
| 56899, | 0.897 AA1: | |
| 56900 | 21 AA2: 22 | MSFEKKILCIGAGYVGGPTMA |
| | Probability: | |
| 603, | 0.916 AA1: | |
| 604 | 30 AA2: 31 | MSFFTWSGISSRSRSFCFGIMMCRALFRLA |
| | Probability: | |
| 77763, | 0.784 AA1: | |
| 77764 | 30 AA2: 31 | MSFICRACSRMTALSRAAPASFRFLNSASA |
| | Probability: | |
| 31877, | 0.845 AA1: | |
| 31878 | 17 AA2: 18 | MSFINTSLGFITAIAFA |
| | Probability: | |
| 8697, | 0.545 AA1: | |
| 8698 | 28 AA2: 29 | MSFIRQPFKYQSFGAVYWLIGINVLVYA |
| | Probability: | |
| 44321, | 1.000 AA1: | |
| 44322 | 27 AA2: 28 | MSFIVKRISSAIITLFLVSVFSFLAFS |
| 45341, | Probability: | |
| 45342 | 1.000 AA1: | MSFKKXFGVSFLFLLFAAASVFG |

| | | |
|---|---|---|
| | 23 AA2: 24 | |
| | Probability: | |
| 14649, | 0.685 AA1: | |
| 14650 | 28 AA2: 29 | MSFLAGAAFGAGWTPCVGPILTGVLLMA |
| | Probability: | |
| 59179, | 0.999 AA1: | |
| 59180 | 27 AA2: 28 | MSFLKKQGRSVCFAVIVIALALMPVFS |
| | Probability: | |
| 20095, | 0.902 AA1: | |
| 20096 | 25 AA2: 26 | MSFPAITWTVSSFFSLNLCNASAMS |
| | Probability: | |
| 35365, | 0.953 AA1: | |
| 35366 | 24 AA2: 25 | MSFRRPTASSICFPPASCSSIALR |
| | Probability: | |
| 48335, | 0.653 AA1: | |
| 48336 | 23 AA2: 24 | MSFSIMLIWFFSKSISTPLKTLA |
| | Probability: | |
| 1185, | 0.988 AA1: | |
| 1186 | 17 AA2: 18 | MSFSLPASLASSSAVLA |
| | Probability: | |
| 11059, | 0.990 AA1: | |
| 11060 | 23 AA2: 24 | MSGERAAVAGGAAVAAIICAAGA |
| | Probability: | |
| 45125, | 0.524 AA1: | |
| 45126 | 16 AA2: 17 | MSGILGSAPLSSTGFA |
| | Probability: | |
| 60005, | 0.823 AA1: | |
| 60006 | 24 AA2: 25 | MSGKKGGFAGMLKLGVVLALYAVA |
| | Probability: | |
| 44323, | 0.995 AA1: | |
| 44324 | 23 AA2: 24 | MSGKLKAGLLLTLLIVLISLTSA |
| | Probability: | |
| 33167, | 0.986 AA1: | |
| 33168 | 23 AA2: 24 | MSGKLKTGLLLTLFIVLISLTSA |

| | | |
|---|---|---|
| | Probability: | |
| 33779, | 0.999 AA1: | |
| 33780 | 28 AA2: 29 | MSGMSNKNRIIIGLFFSLLIIAAAVSLG |
| | Probability: | |
| 35463, | 0.568 AA1: | |
| 35464 | 30 AA2: 31 | MSHLQPSSLKGVKKMTLFSILSVGTRALQT |
| | Probability: | |
| 17085, | 0.623 AA1: | |
| 17086 | 29 AA2: 30 | MSHNTTSPSPRAKSSRATAVPAAPAPSMT |
| | Probability: | |
| 72293, | 0.605 AA1: | |
| 72294 | 18 AA2: 19 | MSIAPSSRSLSNSSSVWA |
| | Probability: | |
| 13573, | 0.754 AA1: | |
| 13574 | 25 AA2: 26 | MSIASTSIRRPVLTIVLSALIVLFG |
| | Probability: | |
| 53131, | 0.899 AA1: | |
| 53132 | 19 AA2: 20 | MSIILTTVIFAAALSLVLG |
| | Probability: | |
| 52077, | 1.000 AA1: | |
| 52078 | 23 AA2: 24 | MSIKKFLGFGFFLLLFAAAFVFG |
| | Probability: | |
| 55535, | 1.000 AA1: | |
| 55536 | 23 AA2: 24 | MSIKRFLGFGFFLLLFAAAFVFG |
| | Probability: | |
| 57535, | 0.983 AA1: | |
| 57536 | 26 AA2: 27 | MSILFTATMIGTPASFAWLIASMVWG |
| | Probability: | |
| 39705, | 0.941 AA1: | |
| 39706 | 26 AA2: 27 | MSILFTATMIGTSASFAWLMASIVCG |
| | Probability: | |
| 43939, | 0.909 AA1: | |
| 43940 | 22 AA2: 23 | MSILKEGYPFIIPAFIAGAAAA |
| 71839, | Probability: | MSILKEGYPFIIPAFIAGAAAA |

| | | |
|---|---|---|
| 71840 | 0.883 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 17519, | 0.542 AA1: | |
| 17520 | 28 AA2: 29 | MSILMVLALPRCESSVFSSFMPSSSVIT |
| | Probability: | |
| 30899, | 0.538 AA1: | |
| 30900 | 28 AA2: 29 | MSILRIGMIVFGILFISIGSGMYFNTHV |
| | Probability: | |
| 14031, | 0.999 AA1: | |
| 14032 | 27 AA2: 28 | MSIMRGRPLPFLFLFLSLIAACSAPVA |
| | Probability: | |
| 33193, | 0.911 AA1: | |
| 33194 | 20 AA2: 21 | MSINGTISSIISITVTSAPS |
| | Probability: | |
| 62009, | 0.972 AA1: | |
| 62010 | 21 AA2: 22 | MSIPLASSTVMTPSLPTLSMA |
| | Probability: | |
| 72037, | 0.791 AA1: | |
| 72038 | 26 AA2: 27 | MSIRIRAFFIIIITNLLIILFSVFLG |
| | Probability: | |
| 16119, | 0.880 AA1: | |
| 16120 | 24 AA2: 25 | MSISVLTSSASVSISVLISHPAAV |
| | Probability: | |
| 28487, | 0.975 AA1: | |
| 28488 | 17 AA2: 18 | MSIVIITFIFATLLAFA |
| | Probability: | |
| 68243, | 0.998 AA1: | |
| 68244 | 19 AA2: 20 | MSIYKKSALALIISSFLFA |
| | Probability: | |
| 60075, | 1.000 AA1: | |
| 60076 | 24 AA2: 25 | MSKFMTVLAATAMFAAVLVGTAAA |
| 70353, | Probability: | |
| 70354 | 0.724 AA1: | MSKFNGKALLPASFVWHWMLPAAIT |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 35965, | 0.798 AA1: | |
| 35966 | 20 AA2: 21 | MSKHKPWALVLSGGGAMGLA |
| | Probability: | |
| 52167, | 1.000 AA1: | |
| 52168 | 27 AA2: 28 | MSKKKAACLLALPPILAAAFSFILPYA |
| | Probability: | |
| 66607, | 0.999 AA1: | |
| 66608 | 21 AA2: 22 | MSKKPFFSFWLVFCLPFIVFA |
| | Probability: | |
| 74041, | 1.000 AA1: | |
| 74042 | 24 AA2: 25 | MSKLFKLLLATAALLTAFSISAKA |
| | Probability: | |
| 71341, | 0.954 AA1: | |
| 71342 | 20 AA2: 21 | MSKLICTGAIDGAVAFVAQA |
| | Probability: | |
| 72087, | 0.954 AA1: | |
| 72088 | 20 AA2: 21 | MSKLICTGAIDGAVAFVAQA |
| | Probability: | |
| 11771, | 0.885 AA1: | |
| 11772 | 20 AA2: 21 | MSKLICTSAIDGAVAIVAHA |
| | Probability: | |
| 61869, | 0.914 AA1: | |
| 61870 | 20 AA2: 21 | MSKLICTSAIDGAVSFVAQA |
| | Probability: | |
| 63019, | 0.911 AA1: | |
| 63020 | 20 AA2: 21 | MSKLICTSAIDGAVSFVAQA |
| | Probability: | |
| 12915, | 0.987 AA1: | |
| 12916 | 26 AA2: 27 | MSKNSIFHLVHLIFIPAFILFAFVFA |
| | Probability: | |
| 76183, | 0.936 AA1: | |
| 76184 | 21 AA2: 22 | MSKNVLFLILAAVVLFFVVQM |

| | | |
|---|---|---|
| | Probability: | |
| 31175, | 0.978 AA1: | |
| 31176 | 27 AA2: 28 | MSKRITIFVISLVALIPALLVVILICA |
| | Probability: | |
| 38425, | 1.000 AA1: | |
| 38426 | 23 AA2: 24 | MSKRVWVLVCTTLLFLCAWSAGA |
| | Probability: | |
| 17577, | 0.581 AA1: | |
| 17578 | 21 AA2: 22 | MSKRYRFAIIIVIVAVCFVFL |
| | Probability: | |
| 16735, | 0.922 AA1: | |
| 16736 | 17 AA2: 18 | MSKSSSASTVSSPPTLS |
| | Probability: | |
| 69867, | 0.999 AA1: | |
| 69868 | 26 AA2: 27 | MSKTLKRFFAVFFITFFVVFFAXSCA |
| | Probability: | |
| 9461, | 0.851 AA1: | |
| 9462 | 22 AA2: 23 | MSKVAAACHCPFVGAATPSFFA |
| | Probability: | |
| 55909, | 0.607 AA1: | |
| 55910 | 18 AA2: 19 | MSKVLIAMSGGVDSSVAA |
| | Probability: | |
| 42805, | 0.940 AA1: | |
| 42806 | 15 AA2: 16 | MSLAMLLYIAFSLLA |
| | Probability: | |
| 47023, | 0.937 AA1: | |
| 47024 | 15 AA2: 16 | MSLAMLLYIAFSLLA |
| | Probability: | |
| 50037, | 0.968 AA1: | |
| 50038 | 23 AA2: 24 | MSLFLPPPLMIVVLSLSILMVLA |
| | Probability: | |
| 39701, | 0.512 AA1: | |
| 39702 | 24 AA2: 25 | MSLKDIKKIIAIASGKGGVGKSTA |
| 19133, | Probability: | MSLKIKLTIIILVMILVSVAVLSVFT |

| | | |
|---|---|---|
| 19134 | 0.998 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| 19777, | 0.988 AA1: | |
| 19778 | 26 AA2: 27 | MSLKVKLTTIILVMILVSIAVLSVFN |
| | Probability: | |
| 10001, | 0.546 AA1: | |
| 10002 | 27 AA2: 28 | MSLLADFHNHSCLSPCGSLDLSPNALA |
| | Probability: | |
| 56333, | 0.885 AA1: | |
| 56334 | 22 AA2: 23 | MSLLPLATSLIAPATRLAPEFA |
| | Probability: | |
| 29327, | 0.504 AA1: | |
| 29328 | 27 AA2: 28 | MSLPPHKWSLAAYRALLGNNGFVNAFA |
| | Probability: | |
| 61399, | 0.998 AA1: | |
| 61400 | 22 AA2: 23 | MSLRIKLTAIILAMILASVAVL |
| | Probability: | |
| 72043, | 0.934 AA1: | |
| 72044 | 24 AA2: 25 | MSLRRPTASSICFPPASCSSIALR |
| | Probability: | |
| 18949, | 0.684 AA1: | |
| 18950 | 23 AA2: 24 | MSLRRQGKIVAFGGRALPGTVQS |
| | Probability: | |
| 26983, | 1.000 AA1: | |
| 26984 | 26 AA2: 27 | MSLSKAGASRLLLGLIILLIVLGAAA |
| | Probability: | |
| 50489, | 1.000 AA1: | |
| 50490 | 26 AA2: 27 | MSLSKTGASRLLLGLIIFLIILGAAA |
| | Probability: | |
| 35315, | 0.990 AA1: | |
| 35316 | 22 AA2: 23 | MSLSNGSIAAIICAAGSSSRMG |
| 53429, | Probability: | |
| 53430 | 0.983 AA1: | MSLSNGSIAAVICAAGSSARMG |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 2239, | 0.626 AA1: | |
| 2240 | 28 AA2: 29 | MSLSNISIGDYGSIAAVICAAGSSARMG |
| | Probability: | |
| 75913, | 0.835 AA1: | |
| 75914 | 23 AA2: 24 | MSLSSMVAAGIAMPVAMRMLANN |
| | Probability: | |
| 46029, | 0.718 AA1: | |
| 46030 | 28 AA2: 29 | MSLTPNETAAILPFAIPAVAACIVPLAA |
| | Probability: | |
| 76179, | 0.956 AA1: | |
| 76180 | 27 AA2: 28 | MSMKKFFVSLFFALLLSCNDGFESKIA |
| | Probability: | |
| 69881, | 0.914 AA1: | |
| 69882 | 21 AA2: 22 | MSMNLWAADAAAVASAAARIA |
| | Probability: | |
| 47995, | 1.000 AA1: | |
| 47996 | 27 AA2: 28 | MSMSLPSIPSFVLSGLLLICLPFSSFA |
| | Probability: | |
| 36701, | 0.845 AA1: | |
| 36702 | 28 AA2: 29 | MSMVKTVVSRPTTFFIIFALLIGLGFFA |
| | Probability: | |
| 64275, | 1.000 AA1: | |
| 64276 | 26 AA2: 27 | MSNFALRSAIFFIAACLWLLTAPVFG |
| | Probability: | |
| 76545, | 0.635 AA1: | |
| 76546 | 27 AA2: 28 | MSNKLDSLEKVLLMGPGPSMVHGSVYA |
| | Probability: | |
| 46963, | 0.910 AA1: | |
| 46964 | 18 AA2: 19 | MSNMAAAATIGSNHNSRA |
| | Probability: | |
| 55147, | 0.910 AA1: | |
| 55148 | 18 AA2: 19 | MSNMAAAATIGSNHNSRA |

| | | |
|---|---|---|
| | Probability: | |
| 10219, | 1.000 AA1: | |
| 10220 | 29 AA2: 30 | MSNNTIILCGMVFSLFLSAMLTMLSAVFS |
| | Probability: | |
| 68111, | 1.000 AA1: | |
| 68112 | 25 AA2: 26 | MSNRKRVLAGLFISILIIIVGASLG |
| | Probability: | |
| 71239, | 0.999 AA1: | |
| 71240 | 25 AA2: 26 | MSNRNRIIIGAVVSLFIIAAAVSLG |
| | Probability: | |
| 76927, | 0.743 AA1: | |
| 76928 | 28 AA2: 29 | MSNVAGILLCFAIALPSWYLGRVFPLMG |
| | Probability: | |
| 44629, | 0.520 AA1: | |
| 44630 | 21 AA2: 22 | MSPAKRLPIITTLAPAPMALV |
| | Probability: | |
| 40771, | 0.917 AA1: | |
| 40772 | 20 AA2: 21 | MSPVKSGRLSALSASCSAIK |
| | Probability: | |
| 19317, | 1.000 AA1: | |
| 19318 | 29 AA2: 30 | MSQKAFLCGFAAFCCLAGLFLLPPAGLYA |
| | Probability: | |
| 11759, | 0.534 AA1: | |
| 11760 | 29 AA2: 30 | MSQPVSFKYLDVVASFFVVILVVSNIASS |
| | Probability: | |
| 68647, | 0.989 AA1: | |
| 68648 | 25 AA2: 26 | MSRFLKIQFPFTIAVLPGLPHSAEA |
| | Probability: | |
| 31823, | 0.633 AA1: | |
| 31824 | 22 AA2: 23 | MSRFRTMIFRALLSYPPYISMA |
| | Probability: | |
| 52475, | 1.000 AA1: | |
| 52476 | 22 AA2: 23 | MSRGKAVGRTLLLTAVMVAALA |
| 9415, | Probability: | MSRKLLDPAVLFFLLLAACSDPSA |

| | | |
|---|---|---|
| 9416 | 0.989 AA1: | |
| | 24 AA2: 25 | |
| | Probability: | |
| 35627, | 0.694 AA1: | |
| 35628 | 26 AA2: 27 | MSRLIHGFRICMSFSLLLSXIGTFMG |
| | Probability: | |
| 13147, | 0.689 AA1: | |
| 13148 | 26 AA2: 27 | MSRNPPNQTSYACGSTVTVTATEASG |
| | Probability: | |
| 59635, | 0.912 AA1: | |
| 59636 | 25 AA2: 26 | MSRQRKTGTVAGMALKLSRASYSGA |
| | Probability: | |
| 7167, | 1.000 AA1: | |
| 7168 | 22 AA2: 23 | MSRRLSIASLFFALVFTLPAQE |
| | Probability: | |
| 23241, | 0.611 AA1: | |
| 23242 | 24 AA2: 25 | MSRRRLMDSSICLPPCSCSSIAFK |
| | Probability: | |
| 5539, | 0.728 AA1: | |
| 5540 | 29 AA2: 30 | MSRSLTPRRMRRQTCMFGILRLCAAINIA |
| | Probability: | |
| 10259, | 0.999 AA1: | |
| 10260 | 21 AA2: 22 | MSRSVLFKIALFVFLSVPAFS |
| | Probability: | |
| 62455, | 0.613 AA1: | |
| 62456 | 26 AA2: 27 | MSRTNAGCRSIPAVMSGHLAFPLTQA |
| | Probability: | |
| 13847, | 1.000 AA1: | |
| 13848 | 25 AA2: 26 | MSSAAPAPAAAPAAAPAAAAPAAAA |
| | Probability: | |
| 36685, | 0.920 AA1: | |
| 36686 | 27 AA2: 28 | MSSALKFRISATLSPFILLIGLLGLMA |
| 44669, | Probability: | |
| 44670 | 0.999 AA1: | MSSFFSTIIITVVSVALLVLISSMA |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 29587, | 0.583 AA1: | |
| 29588 | 18 AA2: 19 | MSSIWTLIWAMRSLISLA |
| | Probability: | |
| 65247, | 0.988 AA1: | |
| 65248 | 28 AA2: 29 | MSSKLGMIGAYVVILVIVFFLGFFAIFA |
| | Probability: | |
| 32527, | 0.982 AA1: | |
| 32528 | 28 AA2: 29 | MSSLASTMRPRAALSGLAFSSIISACKA |
| | Probability: | |
| 58555, | 0.986 AA1: | |
| 58556 | 25 AA2: 26 | MSSLFRCXVLSPAFCLHGGSAPETA |
| | Probability: | |
| 5903, | 0.986 AA1: | |
| 5904 | 28 AA2: 29 | MSSRLGMIGAYIVILVIVFFLGFFAIFA |
| | Probability: | |
| 57935, | 0.883 AA1: | |
| 57936 | 23 AA2: 24 | MSSVLILTIVSIFLVLNSCSRSS |
| | Probability: | |
| 32097, | 1.000 AA1: | |
| 32098 | 19 AA2: 20 | MSSVVIIFLLFFLLTPVFA |
| | Probability: | |
| 24845, | 0.998 AA1: | |
| 24846 | 23 AA2: 24 | MSSVYSVVFIFFLFFAMLPQGFA |
| | Probability: | |
| 18493, | 0.757 AA1: | |
| 18494 | 28 AA2: 29 | MSTAALGKLSCFSISFTVMSPVSRPSSS |
| | Probability: | |
| 23689, | 0.881 AA1: | |
| 23690 | 16 AA2: 17 | MSTADSQLLASASAFA |
| | Probability: | |
| 13563, | 0.688 AA1: | |
| 13564 | 29 AA2: 30 | MSTEYQKIITSLKRRRKGAVAADICAATA |

| | | |
|---|---|---|
| | Probability: | |
| 71003, | 0.998 AA1: | |
| 71004 | 27 AA2: 28 | MSTFIKPLVALIISILLFVGLIFLAHG |
| | Probability: | |
| 8677, | 0.999 AA1: | |
| 8678 | 24 AA2: 25 | MSTFLKPLVALVVSILLFAGLLFL |
| | Probability: | |
| 19301, | 0.922 AA1: | |
| 19302 | 26 AA2: 27 | MSTIRNSKTKTFGILTSGGDCPGLNA |
| | Probability: | |
| 73031, | 1.000 AA1: | |
| 73032 | 25 AA2: 26 | MSTNMRSVALMTLATLIATPNFVLA |
| | Probability: | |
| 44423, | 1.000 AA1: | |
| 44424 | 28 AA2: 29 | MSTPFNPFLPILLLLLLAIGVGAFTLFA |
| | Probability: | |
| 16737, | 1.000 AA1: | |
| 16738 | 29 AA2: 30 | MSTRKHLFMPPAAASILALIFSLATTASA |
| | Probability: | |
| 3625, | 0.954 AA1: | |
| 3626 | 16 AA2: 17 | MSTSALSGGAGAGTFA |
| | Probability: | |
| 51701, | 0.985 AA1: | |
| 51702 | 26 AA2: 27 | MSTSSSKLSASNSAASGGSCSGCAVS |
| | Probability: | |
| 48115, | 0.914 AA1: | |
| 48116 | 26 AA2: 27 | MSTWPITIWKRRSMYFFFSSSLFALS |
| | Probability: | |
| 205, | 0.589 AA1: | |
| 206 | 18 AA2: 19 | MSVASVGFNLVTSMPVHE |
| | Probability: | |
| 14413, | 0.771 AA1: | |
| 14414 | 20 AA2: 21 | MSVFVVVILLAFAACKSTTK |
| 13635, | Probability: | MSVLKEGYSFIFTSFAAGLALVIVSVFM |

| | | |
|---|---|---|
| 13636 | 0.623 AA1: | |
| | 28 AA2: 29 | |
| | Probability: | |
| 18827, | 1.000 AA1: | |
| 18828 | 22 AA2: 23 | MSVRRIFWLLVTLCLVSIPAFS |
| | Probability: | |
| 36271, | 1.000 AA1: | |
| 36272 | 22 AA2: 23 | MSVRRKFLILAILCLISFPAFS |
| | Probability: | |
| 19557, | 0.920 AA1: | |
| 19558 | 22 AA2: 23 | MSVYEGLMDAVTGLSGSGPAFA |
| | Probability: | |
| 39989, | 0.875 AA1: | |
| 39990 | 27 AA2: 28 | MSYAMLLLIGLLTSAHCIAMCGGINLS |
| | Probability: | |
| 17185, | 0.513 AA1: | |
| 17186 | 22 AA2: 23 | MSYGYGLSASPLQVLMAGCAIA |
| | Probability: | |
| 37033, | 0.797 AA1: | |
| 37034 | 28 AA2: 29 | MTAASVAVGAALVGIIAYILIQALGISA |
| | Probability: | |
| 75149, | 0.577 AA1: | |
| 75150 | 19 AA2: 20 | MTAEEAPVSLNLAAALAWA |
| | Probability: | |
| 78695, | 0.819 AA1: | |
| 78696 | 18 AA2: 19 | MTAFDNAISAAKSALASA |
| | Probability: | |
| 48171, | 0.947 AA1: | |
| 48172 | 19 AA2: 20 | MTAFILIIVVLFLCIVLFA |
| | Probability: | |
| 66583, | 0.992 AA1: | |
| 66584 | 20 AA2: 21 | MTAFMFXSFFPVILSALSRA |
| 44783, | Probability: | |
| 44784 | 0.742 AA1: | MTAGSVLAAAVLGVGTVIS |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 27419, | 0.761 AA1: | |
| 27420 | 24 AA2: 25 | MTALLRLGFPTAPGLLSLSLAVHR |
| | Probability: | |
| 69687, | 0.613 AA1: | |
| 69688 | 18 AA2: 19 | MTAQTAVIGAGAWGTAVA |
| | Probability: | |
| 77959, | 0.999 AA1: | |
| 77960 | 22 AA2: 23 | MTDIAKKRTLSLILCAVFAALS |
| | Probability: | |
| 65245, | 0.977 AA1: | |
| 65246 | 21 AA2: 22 | MTDKKFFSFMLTAALGLFALV |
| | Probability: | |
| 72177, | 0.709 AA1: | |
| 72178 | 23 AA2: 24 | MTDKRLGYILMFFVLLLVGGVYG |
| | Probability: | |
| 12011, | 0.986 AA1: | |
| 12012 | 22 AA2: 23 | MTDRQTLLFRLGGFAFVHAVLA |
| | Probability: | |
| 4827, | 0.958 AA1: | |
| 4828 | 29 AA2: 30 | MTDWPSGMASLVLVCMSSMRSDISPAFFA |
| | Probability: | |
| 16255, | 0.942 AA1: | |
| 16256 | 27 AA2: 28 | MTEWPRLLNMPVLSGLLFTGRLPSAFA |
| | Probability: | |
| 68621, | 0.761 AA1: | |
| 68622 | 21 AA2: 22 | MTFANANSASTTFVMPGNAVT |
| | Probability: | |
| 38095, | 0.886 AA1: | |
| 38096 | 14 AA2: 15 | MTFLATAACLSALG |
| | Probability: | |
| 55213, | 0.818 AA1: | |
| 55214 | 24 AA2: 25 | MTFLLSFITLIFIWAIAFPVGIYS |

| | | |
|---|---|---|
| | Probability: | |
| 39541, | 0.615 AA1: | |
| 39542 | 18 AA2: 19 | MTFNCMIAGVGGQGTVLA |
| | Probability: | |
| 47269, | 0.619 AA1: | |
| 47270 | 18 AA2: 19 | MTFNCMIAGVGGQGTVLA |
| | Probability: | |
| 72341, | 0.984 AA1: | |
| 72342 | 21 AA2: 22 | MTFPVLTVLLTSGLTSVLTVA |
| | Probability: | |
| 37743, | 0.600 AA1: | |
| 37744 | 27 AA2: 28 | MTFRAEKCDSFIALGGGSSMDAAKAAA |
| | Probability: | |
| 653, | 1.000 AA1: | |
| 654 | 28 AA2: 29 | MTFRISFKKIILCFLGCMFLPFSLYSLG |
| | Probability: | |
| 58517, | 0.507 AA1: | |
| 58518 | 16 AA2: 17 | MTFSGSSPMLCISSGA |
| | Probability: | |
| 65167, | 0.509 AA1: | |
| 65168 | 16 AA2: 17 | MTFSGSSPMLCISSGA |
| | Probability: | |
| 68365, | 0.652 AA1: | |
| 68366 | 22 AA2: 23 | MTFSLRSSRAFWTSALSLSAIL |
| | Probability: | |
| 29443, | 0.861 AA1: | |
| 29444 | 25 AA2: 26 | MTGGGDLSANRTLAISAATSATASA |
| | Probability: | |
| 52419, | 1.000 AA1: | |
| 52420 | 26 AA2: 27 | MTGINAAALAFASAAALALAAALALA |
| | Probability: | |
| 8653, | 0.657 AA1: | |
| 8654 | 30 AA2: 31 | MTGIRIAARAELSRRNEMLLRLFAAISALA |
| 52243, | Probability: | MTGLSINFPFSYKIFLPVFLLLCLLFA |

| | | |
|---|---|---|
| 52244 | 0.713 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| 78425, | 0.721 AA1: | |
| 78426 | 17 AA2: 18 | MTICPFSSAIPSTPAPA |
| | Probability: | |
| 58351, | 0.872 AA1: | |
| 58352 | 26 AA2: 27 | MTIFIKNRPKRFSKKASCSECSPAQA |
| | Probability: | |
| 59241, | 1.000 AA1: | |
| 59242 | 23 AA2: 24 | MTIFKRSLLTLGAAALVFSAVLV |
| | Probability: | |
| 64747, | 0.999 AA1: | |
| 64748 | 23 AA2: 24 | MTIFKRSLLTLGTAALVFSAVLV |
| | Probability: | |
| 47913, | 0.939 AA1: | |
| 47914 | 23 AA2: 24 | MTIIIITAIFALVLAFVIGTALG |
| | Probability: | |
| 34125, | 1.000 AA1: | |
| 34126 | 21 AA2: 22 | MTIIRAAAIALLASLSLTAQV |
| | Probability: | |
| 54411, | 0.740 AA1: | |
| 54412 | 22 AA2: 23 | MTIIRARVLGFCMGVRRAAELA |
| | Probability: | |
| 60109, | 0.996 AA1: | |
| 60110 | 19 AA2: 20 | MTIKVKLSLMMIAIMVVVA |
| | Probability: | |
| 20945, | 0.965 AA1: | |
| 20946 | 21 AA2: 22 | MTIQRDTKVLFAIAFAVIAVA |
| | Probability: | |
| 39795, | 0.644 AA1: | |
| 39796 | 20 AA2: 21 | MTISRITPSTPSMICLTTFS |
| 14211, | Probability: | |
| 14212 | 0.905 AA1: | MTISSPPGRRTLISVSLSQCPFRTA |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 47263, | 0.998 AA1: | |
| 47264 | 30 AA2: 31 | MTKFRLYLILGALFLVLGCSAKYGSSNYAS |
| | Probability: | |
| 74417, | 0.917 AA1: | |
| 74418 | 24 AA2: 25 | MTKFSKKKLFIYSNLFAVPPSLLA |
| | Probability: | |
| 61615, | 0.998 AA1: | |
| 61616 | 26 AA2: 27 | MTKFTILLAIVSIIIAAVIIFFTLQS |
| | Probability: | |
| 51753, | 0.978 AA1: | |
| 51754 | 23 AA2: 24 | MTKHFFSFIFFFFLFYSILPAQE |
| | Probability: | |
| 10005, | 1.000 AA1: | |
| 10006 | 26 AA2: 27 | MTKIKFYKHILACFLFFLLFFSCATS |
| | Probability: | |
| 4677, | 0.999 AA1: | |
| 4678 | 30 AA2: 31 | MTKKHKALAAVIACVLFWGFSFISIKIAVA |
| | Probability: | |
| 41755, | 1.000 AA1: | |
| 41756 | 25 AA2: 26 | MTKKITMIWVLLFFAGFCYGQSSFA |
| | Probability: | |
| 5987, | 1.000 AA1: | |
| 5988 | 17 AA2: 18 | MTKKNAAFLLALVPALA |
| | Probability: | |
| 71423, | 0.977 AA1: | |
| 71424 | 25 AA2: 26 | MTKKQVTIITALSVIAFILALMVSG |
| | Probability: | |
| 29781, | 0.723 AA1: | |
| 29782 | 28 AA2: 29 | MTKKRLIIRLDRTPGAGTILGATGGVME |
| | Probability: | |
| 52739, | 1.000 AA1: | |
| 52740 | 17 AA2: 18 | MTKKTAACLLALPPVLA |

| | | |
|---|---|---|
| | Probability: | |
| 13907, | 0.545 AA1: | |
| 13908 | 25 AA2: 26 | MTKLSSYKNYVPFLGFISLAALFQG |
| | Probability: | |
| 12031, | 1.000 AA1: | |
| 12032 | 25 AA2: 26 | MTKNRVNFVKLFLTAVLALVTFAMG |
| | Probability: | |
| 49157, | 0.999 AA1: | |
| 49158 | 26 AA2: 27 | MTKRFSLHSTIYVCAALALGAFFSQS |
| | Probability: | |
| 5869, | 0.922 AA1: | |
| 5870 | 27 AA2: 28 | MTKVVNKFCIFTGALVLIVAGFSPKIG |
| | Probability: | |
| 73257, | 0.995 AA1: | |
| 73258 | 26 AA2: 27 | MTKWFKRNWIWMLVLGAAVVYAPLVG |
| | Probability: | |
| 20233, | 0.950 AA1: | |
| 20234 | 25 AA2: 26 | MTLAISATPRLICSIPLACSPEAWA |
| | Probability: | |
| 44879, | 0.814 AA1: | |
| 44880 | 20 AA2: 21 | MTLATLVIIALSLLSLLKFA |
| | Probability: | |
| 65157, | 0.990 AA1: | |
| 65158 | 23 AA2: 24 | MTLERPILAIAALVIFPIAIALA |
| | Probability: | |
| 14177, | 0.936 AA1: | |
| 14178 | 23 AA2: 24 | MTLERPVLAIATLVIFPLAIALA |
| | Probability: | |
| 60413, | 0.985 AA1: | |
| 60414 | 24 AA2: 25 | MTLKPLILAPLLLALCCQTPQVRT |
| | Probability: | |
| 74357, | 0.637 AA1: | |
| 74358 | 21 AA2: 22 | MTLLRWLLAPLAPLYGLAVRA |
| 51151, | Probability: | MTLLSSVLLGGFALALGLSLLG |

| | | |
|---|---|---|
| 51152 | 0.997 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 72715, | 0.820 AA1: | |
| 72716 | 25 AA2: 26 | MTLLTKFLPIGLLALGIIAGVQTWR |
| | Probability: | |
| 9693, | 0.993 AA1: | |
| 9694 | 27 AA2: 28 | MTLSSVLPALLLLLSGPXFIDRALAQA |
| | Probability: | |
| 18523, | 0.787 AA1: | |
| 18524 | 27 AA2: 28 | MTLSTLIAFAISIGLFMYAPMWTAAHV |
| | Probability: | |
| 14767, | 0.998 AA1: | |
| 14768 | 21 AA2: 22 | MTLVLLTVVLISGLTSVLTLA |
| | Probability: | |
| 65303, | 0.976 AA1: | |
| 65304 | 25 AA2: 26 | MTMGAQFESFLLPLVLMLSIPFSLA |
| | Probability: | |
| 26567, | 1.000 AA1: | |
| 26568 | 19 AA2: 20 | MTMKKKGLFAILAVLAVFA |
| | Probability: | |
| 74063, | 0.998 AA1: | |
| 74064 | 22 AA2: 23 | MTMTIKLFAAIAAVVLLGACSS |
| | Probability: | |
| 67575, | 0.991 AA1: | |
| 67576 | 17 AA2: 18 | MTNGAFPLSGAPAAAGA |
| | Probability: | |
| 38463, | 0.783 AA1: | |
| 38464 | 27 AA2: 28 | MTNIILDSREAGAFITTSVSAIAPANN |
| | Probability: | |
| 25287, | 0.886 AA1: | |
| 25288 | 18 AA2: 19 | MTNKALIVGAGGVANVVA |
| 19463, | Probability: | |
| 19464 | 0.987 AA1: | MTNLNLLFIPALIAAVLLA |

| | | |
|---|---|---|
| | 19 AA2: 20 | |
| | Probability: | |
| 4275, | 0.771 AA1: | |
| 4276 | 29 AA2: 30 | MTNLRGLLSKNIKKRRKFLGISQAILAEK |
| | Probability: | |
| 29069, | 0.996 AA1: | |
| 29070 | 28 AA2: 29 | MTNMSNNTIILCGMVLSLFLSAILTTIS |
| | Probability: | |
| 49083, | 0.953 AA1: | |
| 49084 | 24 AA2: 25 | MTNPTKTSRGGFVSILLLMSVVLG |
| | Probability: | |
| 22329, | 0.629 AA1: | |
| 22330 | 29 AA2: 30 | MTPFVIYMASESLGVSGILAVFASGIAHS |
| | Probability: | |
| 60351, | 0.508 AA1: | |
| 60352 | 25 AA2: 26 | MTPGRIFVFSAPSGAGKSTILSRLA |
| | Probability: | |
| 14703, | 0.847 AA1: | |
| 14704 | 28 AA2: 29 | MTQMSASAKRPFISAAISAAVLTRMTFA |
| | Probability: | |
| 51937, | 0.937 AA1: | |
| 51938 | 26 AA2: 27 | MTRALASFAALGSVWYFALLGFPAFS |
| | Probability: | |
| 33525, | 0.515 AA1: | |
| 33526 | 28 AA2: 29 | MTRERAFRYDFFMMAAILLLLMLGLVIL |
| | Probability: | |
| 49925, | 0.965 AA1: | |
| 49926 | 18 AA2: 19 | MTRKQASTISIMILAVVA |
| | Probability: | |
| 8087, | 0.888 AA1: | |
| 8088 | 28 AA2: 29 | MTRNKTITLLILEVIFLALLFMVWNSFR |
| | Probability: | |
| 64041, | 0.948 AA1: | |
| 64042 | 19 AA2: 20 | MTRSGKGVSTVMTTLLAYA |

| | | |
|---|---|---|
| | Probability: | |
| 33347, | 0.522 AA1: | |
| 33348 | 20 AA2: 21 | MTSANVPVIAVSAGNSHSMA |
| | Probability: | |
| 78309, | 0.997 AA1: | |
| 78310 | 25 AA2: 26 | MTSESLSRRQFIGTGAALGGSSALA |
| | Probability: | |
| 30441, | 0.996 AA1: | |
| 30442 | 21 AA2: 22 | MTSEYWILLVFLFLGAFIAGA |
| | Probability: | |
| 74697, | 0.875 AA1: | |
| 74698 | 25 AA2: 26 | MTSFLYXVVSCTAIIIIASXGGLYA |
| | Probability: | |
| 21889, | 0.548 AA1: | |
| 21890 | 22 AA2: 23 | MTSKDQKMLSAAAIFSNNMVLQ |
| | Probability: | |
| 40367, | 0.887 AA1: | |
| 40368 | 29 AA2: 30 | MTSLIFLMTSLIFLSTSLKFLMTSLIFLS |
| | Probability: | |
| 73581, | 0.801 AA1: | |
| 73582 | 23 AA2: 24 | MTSLKKFLTIAGSDCCGGAGVQA |
| | Probability: | |
| 72339, | 0.995 AA1: | |
| 72340 | 17 AA2: 18 | MTSPSAGLAAFLPCANA |
| | Probability: | |
| 273, | 0.726 AA1: | |
| 274 | 21 AA2: 22 | MTSRPRSPISATTFTSALVLR |
| | Probability: | |
| 36317, | 0.729 AA1: | |
| 36318 | 22 AA2: 23 | MTSSSISGTTTRPMTASISSSA |
| | Probability: | |
| 13461, | 0.999 AA1: | |
| 13462 | 23 AA2: 24 | MTSVLTVTLTLGLTFAVSVLTLA |
| 14137, | Probability: | MTTFKSGRLYIFGAGFAGQTLA |

| | | |
|---|---|---|
| 14138 | 0.943 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 31953, | 0.943 AA1: | |
| 31954 | 22 AA2: 23 | MTTFKSGRLYIFGAGFAGQTLA |
| | Probability: | |
| 76999, | 0.855 AA1: | |
| 77000 | 20 AA2: 21 | MTTKSLGSTTLLYLATQVFT |
| | Probability: | |
| 34389, | 0.806 AA1: | |
| 34390 | 26 AA2: 27 | MTTPNPGITAAGIAIXAVLGXLSVLS |
| | Probability: | |
| 59393, | 0.895 AA1: | |
| 59394 | 21 AA2: 22 | MTTQGDVVSAGSIFVTTNLAA |
| | Probability: | |
| 64209, | 0.973 AA1: | |
| 64210 | 21 AA2: 22 | MTTSLILAASSLETPSLSGFG |
| | Probability: | |
| 64447, | 0.997 AA1: | |
| 64448 | 23 AA2: 24 | MTTSPLKITLALAVAIITSCSSS |
| | Probability: | |
| 77461, | 0.517 AA1: | |
| 77462 | 21 AA2: 22 | MTTSVCTACPNRASRSPTALA |
| | Probability: | |
| 3211, | 0.975 AA1: | |
| 3212 | 24 AA2: 25 | MTTVFLALVXTWRISVIMPFISLA |
| | Probability: | |
| 65243, | 1.000 AA1: | |
| 65244 | 24 AA2: 25 | MTTYKKNRFYSVLVFSLFILPAIA |
| | Probability: | |
| 60897, | 0.540 AA1: | |
| 60898 | 18 AA2: 19 | MTVFGGTGATGSGNYAQG |
| 10833, | Probability: | |
| 10834 | 0.902 AA1: | MTVKFPYRSTIILSVTGILAVA |

| | | |
|---|---|---|
| | 22 AA2: 23 | |
| | Probability: | |
| 49641, | 0.598 AA1: | |
| 49642 | 19 AA2: 20 | MTVLFLLMMAAIAGILVIS |
| | Probability: | |
| 27967, | 1.000 AA1: | |
| 27968 | 25 AA2: 26 | MTVSRKAALSLLISVLLFAVFIVLA |
| | Probability: | |
| 76679, | 0.999 AA1: | |
| 76680 | 25 AA2: 26 | MTVTAPSSSSVAASSSAGQSSSSVA |
| | Probability: | |
| 41925, | 0.852 AA1: | |
| 41926 | 22 AA2: 23 | MTWSEWPSSAIAALSCATKLRA |
| | Probability: | |
| 8463, | 0.995 AA1: | |
| 8464 | 21 AA2: 22 | MTWSSSSTSVATVTSAGLVTA |
| | Probability: | |
| 28349, | 0.871 AA1: | |
| 28350 | 23 AA2: 24 | MTYKGVFREFLLFFAGFSLFAQG |
| | Probability: | |
| 63339, | 0.937 AA1: | |
| 63340 | 28 AA2: 29 | MVAEFDLPFAFPSVAFLFCVLAASASSK |
| | Probability: | |
| 16773, | 0.533 AA1: | |
| 16774 | 23 AA2: 24 | MVANGNLAAKVVLVVLLSFSVFS |
| | Probability: | |
| 37289, | 0.894 AA1: | |
| 37290 | 28 AA2: 29 | MVANMPLCSSFATTSLAGRLHFSASSLT |
| | Probability: | |
| 12327, | 0.867 AA1: | |
| 12328 | 21 AA2: 22 | MVANRLAGLANTLAAFPNRLA |
| | Probability: | |
| 66375, | 0.872 AA1: | |
| 66376 | 21 AA2: 22 | MVAVILMAVFLYVFAALPDSA |

| | | |
|---|---|---|
| | Probability: | |
| 2735, | 0.517 AA1: | |
| 2736 | 18 AA2: 19 | MVEELTFSAXAPVATAKA |
| | Probability: | |
| 58429, | 0.557 AA1: | |
| 58430 | 26 AA2: 27 | MVENAVLQNILLFGPVFLLIAVRALA |
| | Probability: | |
| 31693, | 0.839 AA1: | |
| 31694 | 30 AA2: 31 | MVFKFMLLMSLSNIIWGVLTCAWFGFDVSA |
| | Probability: | |
| 52725, | 0.995 AA1: | |
| 52726 | 19 AA2: 20 | MVFKYLVLATLALAISAYS |
| | Probability: | |
| 6339, | 0.505 AA1: | |
| 6340 | 21 AA2: 22 | MVFLLNFYSALFIVLTSPYQQ |
| | Probability: | |
| 20467, | 0.723 AA1: | |
| 20468 | 21 AA2: 22 | MVFVTAGMGGGTGTGAAPVIA |
| | Probability: | |
| 36273, | 0.667 AA1: | |
| 36274 | 24 AA2: 25 | MVGLSWAVVAALVALFMFVSRTFR |
| | Probability: | |
| 11537, | 0.705 AA1: | |
| 11538 | 27 AA2: 28 | MVGMMLCFSMTFWFVFQQAGGTLNLFA |
| | Probability: | |
| 14947, | 0.539 AA1: | |
| 14948 | 30 AA2: 31 | MVIGLVRKLIKPVFLAALTTMAGFISFCFT |
| | Probability: | |
| 8229, | 0.672 AA1: | |
| 8230 | 25 AA2: 26 | MVIISTLLKITGGLCLFLYGMKVMS |
| | Probability: | |
| 31739, | 0.684 AA1: | |
| 31740 | 27 AA2: 28 | MVIKVVLLLLFFAVTISVGFYFRRKVA |
| 525, | Probability: | MVILKINHNKRHNWRCFCAGLIVMQAIFA |

| | | |
|---|---|---|
| 526 | 0.994 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| 36511, | 0.940 AA1: | |
| 36512 | 19 AA2: 20 | MVIMILAAVIFSAIFSLFN |
| | Probability: | |
| 71555, | 0.997 AA1: | |
| 71556 | 23 AA2: 24 | MVINNVIPIVFILIPVMASSVMA |
| | Probability: | |
| 60765, | 0.721 AA1: | |
| 60766 | 20 AA2: 21 | MVIPATASCFITANTSPTSS |
| | Probability: | |
| 42937, | 0.509 AA1: | |
| 42938 | 24 AA2: 25 | MVIPSRASCFITANTSPTSSGSRA |
| | Probability: | |
| 58967, | 0.859 AA1: | |
| 58968 | 24 AA2: 25 | MVIPSRASCFITANTSPTSSGSSA |
| | Probability: | |
| 66901, | 0.976 AA1: | |
| 66902 | 27 AA2: 28 | MVIRFSFITMLLLIFSVFWLSACTGMK |
| | Probability: | |
| 43845, | 0.998 AA1: | |
| 43846 | 26 AA2: 27 | MVKAARLFFKVLSVLIVLALVLFSLA |
| | Probability: | |
| 22117, | 0.915 AA1: | |
| 22118 | 18 AA2: 19 | MVKAFVAIIFVVVSIANS |
| | Probability: | |
| 22749, | 0.998 AA1: | |
| 22750 | 24 AA2: 25 | MVKIFTRFIFICIVAIFLVFGVMA |
| | Probability: | |
| 17027, | 0.994 AA1: | |
| 17028 | 18 AA2: 19 | MVKKIIFALLLTIGAALL |
| 29275, | Probability: | |
| 29276 | 0.997 AA1: | MVKKLITGLAFSLMTMLLFS |

| | | |
|---|---|---|
| | 20 AA2: 21 | |
| | Probability: | |
| 77261, | 0.707 AA1: | |
| 77262 | 24 AA2: 25 | MVKRIRPKVTVPLLFMTYINPVFT |
| | Probability: | |
| 56213, | 0.561 AA1: | |
| 56214 | 22 AA2: 23 | MVKTLRKTIPNLTLVGGNVATA |
| | Probability: | |
| 40147, | 0.550 AA1: | |
| 40148 | 21 AA2: 22 | MVLCLAMIGVCGTGGQGSLHG |
| | Probability: | |
| 421, | 0.657 AA1: | |
| 422 | 29 AA2: 30 | MVLHCRRIQMKISFKLTAIMFVLSLFSIG |
| | Probability: | |
| 4507, | 0.846 AA1: | |
| 4508 | 22 AA2: 23 | MVLKFFLLFGAFLLISCGTIER |
| | Probability: | |
| 61659, | 0.834 AA1: | |
| 61660 | 24 AA2: 25 | MVLKKWNRVFIFFFFALILVSCKS |
| | Probability: | |
| 13355, | 0.715 AA1: | |
| 13356 | 22 AA2: 23 | MVLLLTMPIAISSAIIAEIVAA |
| | Probability: | |
| 7843, | 1.000 AA1: | |
| 7844 | 16 AA2: 17 | MVLPAILLLLCASALA |
| | Probability: | |
| 261, | 0.961 AA1: | |
| 262 | 19 AA2: 20 | MVLRFFLFASLFLFASCYE |
| | Probability: | |
| 34941, | 0.990 AA1: | |
| 34942 | 24 AA2: 25 | MVLWNKLGLCFILLIMLFSFSKCS |
| | Probability: | |
| 26597, | 0.673 AA1: | |
| 26598 | 27 AA2: 28 | MVLYIIKRAAFVVITLLFIICLTYILM |

| | | |
|---|---|---|
| | Probability: | |
| 60121, | 0.869 AA1: | |
| 60122 | 19 AA2: 20 | MVMIIASVFTSCISPAENS |
| | Probability: | |
| 42389, | 0.986 AA1: | |
| 42390 | 19 AA2: 20 | MVMKTSTAALNFAAVMAQA |
| | Probability: | |
| 39579, | 0.964 AA1: | |
| 39580 | 22 AA2: 23 | MVMQILFTFVQALGICSLLALA |
| | Probability: | |
| 8287, | 0.851 AA1: | |
| 8288 | 25 AA2: 26 | MVNMRFIYASIIFLLLFPSCTKREA |
| | Probability: | |
| 70967, | 0.817 AA1: | |
| 70968 | 21 AA2: 22 | MVNMRIIFVFIICLLFLPACT |
| | Probability: | |
| 43897, | 0.971 AA1: | |
| 43898 | 29 AA2: 30 | MVNMRLAGKTALISGAAQGLGAAMAYRFA |
| | Probability: | |
| 65197, | 0.996 AA1: | |
| 65198 | 20 AA2: 21 | MVQGLLAFVMAAPVARAAGA |
| | Probability: | |
| 78223, | 0.515 AA1: | |
| 78224 | 18 AA2: 19 | MVRIGRGGSLTILATALT |
| | Probability: | |
| 31295, | 0.954 AA1: | |
| 31296 | 25 AA2: 26 | MVRKESVGKKVLNLFSYTGSFSVYA |
| | Probability: | |
| 51079, | 0.990 AA1: | |
| 51080 | 21 AA2: 22 | MVRLAEILVVLALCGVFSSCA |
| | Probability: | |
| 40595, | 0.951 AA1: | |
| 40596 | 28 AA2: 29 | MVRVFRGKFFSFLFFITLISLSSCSKLG |
| 54065, | Probability: | MVSAILLMIIYLSFISLG |

| | | |
|---|---|---|
| 54066 | 0.612 AA1: | |
| | 18 AA2: 19 | |
| | Probability: | |
| 51211, | 0.508 AA1: | |
| 51212 | 22 AA2: 23 | MVSEVDVAVIGGGPAGLAAAIA |
| | Probability: | |
| 41685, | 0.977 AA1: | |
| 41686 | 21 AA2: 22 | MVSFFLVLLLVSIMLIRSRTA |
| | Probability: | |
| 33063, | 0.616 AA1: | |
| 33064 | 22 AA2: 23 | MVSIAFTGGGTGGHIYPGLAVA |
| | Probability: | |
| 71475, | 0.622 AA1: | |
| 71476 | 22 AA2: 23 | MVSIAFTGGGTGGHIYPGLAVA |
| | Probability: | |
| 77269, | 1.000 AA1: | |
| 77270 | 25 AA2: 26 | MVSITSGAAAALPALISAIISSSMS |
| | Probability: | |
| 58339, | 0.736 AA1: | |
| 58340 | 17 AA2: 18 | MVSLKSFLSSALSIVSG |
| | Probability: | |
| 25221, | 0.903 AA1: | |
| 25222 | 21 AA2: 22 | MVSLLILAAAYFASLLMHLMV |
| | Probability: | |
| 71879, | 0.555 AA1: | |
| 71880 | 17 AA2: 18 | MVSLNIFLSSALSIVSG |
| | Probability: | |
| 15165, | 0.613 AA1: | |
| 15166 | 19 AA2: 20 | MVSSICFLICATRAFISLG |
| | Probability: | |
| 16123, | 0.998 AA1: | |
| 16124 | 29 AA2: 30 | MVSSSSTNLSSSVVATSSSSRATSSSSRA |
| 8827, | Probability: | |
| 8828 | 0.998 AA1: | MVSSTSSLSSVSSISLVSSSSSTS |

| | | |
|---|---|---|
| | 24 AA2: 25 | |
| | Probability: | |
| 16677, | 0.932 AA1: | |
| 16678 | 26 AA2: 27 | MVTHVFLAGLFFFLMLRRSFGVARLA |
| | Probability: | |
| 49747, | 0.713 AA1: | |
| 49748 | 19 AA2: 20 | MVTLLLCFFAVLFNPVTDS |
| | Probability: | |
| 303, | 0.828 AA1: | |
| 304 | 25 AA2: 26 | MVVSMGAVAASGGYWASMSASRIVA |
| | Probability: | |
| 58097, | 0.518 AA1: | |
| 58098 | 24 AA2: 25 | MVWLYMIILMAGITTIPISLYESA |
| | Probability: | |
| 34293, | 0.946 AA1: | |
| 34294 | 23 AA2: 24 | MVWRMNKKALIAMSGGVDSSVAA |
| | Probability: | |
| 77071, | 0.724 AA1: | |
| 77072 | 21 AA2: 22 | MVWYSSITSSATSGITCLVSA |
| | Probability: | |
| 28951, | 0.732 AA1: | |
| 28952 | 24 AA2: 25 | MVYQGEKLSASLSASWLAGGGLSG |
| | Probability: | |
| 25583, | 0.998 AA1: | |
| 25584 | 26 AA2: 27 | MVYRCNIKSFVLMAALFGLFSYSIFA |
| | Probability: | |
| 15927, | 0.847 AA1: | |
| 15928 | 27 AA2: 28 | MWAIILLSAFILILIISNVRIVSQSNA |
| | Probability: | |
| 45605, | 0.855 AA1: | |
| 45606 | 20 AA2: 21 | MWAVYALFSALFAAMTSILG |
| | Probability: | |
| 45175, | 0.979 AA1: | |
| 45176 | 27 AA2: 28 | MWDFRLPRALAATLAGASLSLSGLLLQ |

| | | |
|---|---|---|
| | Probability: | |
| 6843, | 0.979 AA1: | |
| 6844 | 27 AA2: 28 | MWDFRXPRALAATLAGASLSLSGLLLQ |
| | Probability: | |
| 53243, | 0.997 AA1: | |
| 53244 | 20 AA2: 21 | MWFIFALLSAFFAALTSILA |
| | Probability: | |
| 75855, | 0.978 AA1: | |
| 75856 | 26 AA2: 27 | MWHITLAGIKPTIIILLIFAAGGLLA |
| | Probability: | |
| 66691, | 0.997 AA1: | |
| 66692 | 20 AA2: 21 | MWIIFALLSAFFAALTSILA |
| | Probability: | |
| 71091, | 0.997 AA1: | |
| 71092 | 20 AA2: 21 | MWIIFALLSAFFAALTSILA |
| | Probability: | |
| 51095, | 0.999 AA1: | |
| 51096 | 20 AA2: 21 | MWIIFALLSALFAALTSILA |
| | Probability: | |
| 4921, | 0.648 AA1: | |
| 4922 | 20 AA2: 21 | MWIKAKNNVGVSGFSPAASG |
| | Probability: | |
| 38415, | 0.613 AA1: | |
| 38416 | 26 AA2: 27 | MWKCEMTNCKMTVILAVFMLFFPLFS |
| | Probability: | |
| 78615, | 0.862 AA1: | |
| 78616 | 24 AA2: 25 | MWKPKRPTRTRPLAYVCAFLSLSA |
| | Probability: | |
| 35991, | 0.535 AA1: | |
| 35992 | 18 AA2: 19 | MWLINLCAFMVVIQSSLT |
| | Probability: | |
| 25181, | 0.994 AA1: | |
| 25182 | 23 AA2: 24 | MWLLRLPRVLLAFFAGGCLAVSG |
| 36615, | Probability: | MWLVAVCAFIATLQSAVG |

| | | |
|---|---|---|
| 36616 | 0.936 AA1: | |
| | 18 AA2: 19 | |
| | Probability: | |
| 61391, | 0.973 AA1: | |
| 61392 | 18 AA2: 19 | MWLVCACAFLCVIQSAFN |
| | Probability: | |
| 73475, | 0.997 AA1: | |
| 73476 | 20 AA2: 21 | MWNRLFILATALLFSSCAEK |
| | Probability: | |
| 74775, | 0.834 AA1: | |
| 74776 | 29 AA2: 30 | MWNSFIKLLEAASAIFFCASLDSTAPPMA |
| | Probability: | |
| 63445, | 0.539 AA1: | |
| 63446 | 24 AA2: 25 | MWRPNTGFPLSPALPAFPRLPTWA |
| | Probability: | |
| 53709, | 0.587 AA1: | |
| 53710 | 27 AA2: 28 | MWRRPASICIVSFTFGGSTIIVWKRLA |
| | Probability: | |
| 29159, | 0.852 AA1: | |
| 29160 | 25 AA2: 26 | MWSETRAACCILCVTMMMVSWLFNS |
| | Probability: | |
| 71479, | 0.648 AA1: | |
| 71480 | 23 AA2: 24 | MWSIIVISAFVLILIISNVRIVS |
| | Probability: | |
| 42579, | 0.986 AA1: | |
| 42580 | 27 AA2: 28 | MWTLSKTLRARRLASFSALSTPMYSRT |
| | Probability: | |
| 76719, | 0.880 AA1: | |
| 76720 | 19 AA2: 20 | MWTLTASAGAAATPFQASR |
| | Probability: | |
| 3817, | 0.687 AA1: | |
| 3818 | 19 AA2: 20 | MWVWSALLPSPVFFIAVRG |
| 34715, | Probability: | |
| 34716 | 0.721 AA1: | MXCFIXSLVSWLRLSPLSSISPALS |

| | | |
|---|---|---|
| | 25 AA2: 26 | |
| | Probability: | |
| 13055, | 0.787 AA1: | |
| 13056 | 19 AA2: 20 | MXGFRFTSFFFLVSGMAFR |
| | Probability: | |
| 7907, | 0.626 AA1: | |
| 7908 | 22 AA2: 23 | MXHARLIPVNLSSSTISSAAMV |
| | Probability: | |
| 46705, | 0.533 AA1: | |
| 46706 | 21 AA2: 22 | MXINKEAAVSAFNRGISAAAA |
| | Probability: | |
| 60767, | 0.744 AA1: | |
| 60768 | 26 AA2: 27 | MXKGARVTISLTVLAVGAGLVLSLFL |
| | Probability: | |
| 13165, | 0.632 AA1: | |
| 13166 | 22 AA2: 23 | MXLAIVISMASLPLTSCNNGNG |
| | Probability: | |
| 52165, | 0.610 AA1: | |
| 52166 | 23 AA2: 24 | MYAKWKAIPKKMVAVAFSANGGS |
| | Probability: | |
| 68835, | 0.976 AA1: | |
| 68836 | 17 AA2: 18 | MYDKLLVFSVLLFSVFA |
| | Probability: | |
| 6289, | 0.563 AA1: | |
| 6290 | 27 AA2: 28 | MYFIPGRFTIVMISLVSISITVSTKLA |
| | Probability: | |
| 28345, | 0.629 AA1: | |
| 28346 | 18 AA2: 19 | MYFLTGGAGFIGSNIAAA |
| | Probability: | |
| 70171, | 0.976 AA1: | |
| 70172 | 27 AA2: 28 | MYGIKFYKHILVCFLFLSLIFSCSTSK |
| | Probability: | |
| 58921, | 0.749 AA1: | |
| 58922 | 19 AA2: 20 | MYIFNKTWIYAFVGVFALA |

| | | |
|---|---|---|
| | Probability: | |
| 28923, | 0.978 AA1: | |
| 28924 | 25 AA2: 26 | MYKINRPSRNRLVFAAVFAALIAVS |
| | Probability: | |
| 47041, | 0.680 AA1: | |
| 47042 | 26 AA2: 27 | MYKISYYSLRGIFFLFVSILFLSCAT |
| | Probability: | |
| 24733, | 0.985 AA1: | |
| 24734 | 25 AA2: 26 | MYKNFPLSFHTSAPAAAAAAIAAHK |
| | Probability: | |
| 64519, | 0.999 AA1: | |
| 64520 | 21 AA2: 22 | MYKRLLLVLFAGLLLNVHVFA |
| | Probability: | |
| 69715, | 1.000 AA1: | |
| 69716 | 19 AA2: 20 | MYKRVFIIILIFFAAVVFA |
| | Probability: | |
| 12103, | 0.949 AA1: | |
| 12104 | 22 AA2: 23 | MYKTFVYVMSACLAGIAGAVFV |
| | Probability: | |
| 12131, | 0.734 AA1: | |
| 12132 | 29 AA2: 30 | MYKWYALKRILRGVVMYALLILLFSVLFN |
| | Probability: | |
| 2681, | 0.934 AA1: | |
| 2682 | 26 AA2: 27 | MYLMEMKFKKMLTVVLTAAVICGFSA |
| | Probability: | |
| 38181, | 1.000 AA1: | |
| 38182 | 27 AA2: 28 | MYMSNRKRVIAGFIISLLLIMAAASLG |
| | Probability: | |
| 6593, | 0.873 AA1: | |
| 6594 | 26 AA2: 27 | MYNASTGNITISGGTVSVATSSNVIA |
| | Probability: | |
| 48973, | 0.953 AA1: | |
| 48974 | 18 AA2: 19 | MYNVSSSAVTISGGSVSA |
| 63149, | Probability: | MYPELLAPAGSFACAEAAFTHG |

| | | |
|---|---|---|
| 63150 | 0.939 AA1: | |
| | 22 AA2: 23 | |
| | Probability: | |
| 29843, | 0.976 AA1: | |
| 29844 | 23 AA2: 24 | MYPIIYTALLAVMMSGCYSFSGA |
| | Probability: | |
| 52239, | 0.622 AA1: | |
| 52240 | 22 AA2: 23 | MYPLLALPIAYLLGSIPTAVWV |
| | Probability: | |
| 10919, | 0.901 AA1: | |
| 10920 | 27 AA2: 28 | MYQLGYKAVGIIALLSFLIGVVLSLQS |
| | Probability: | |
| 42089, | 0.636 AA1: | |
| 42090 | 22 AA2: 23 | MYQSIAILISGFFCGIGGSFLS |
| | Probability: | |
| 35765, | 0.592 AA1: | |
| 35766 | 28 AA2: 29 | MYRRFYLDNSAPVLLVLALFCLFNNVGA |
| | Probability: | |
| 22909, | 0.920 AA1: | |
| 22910 | 25 AA2: 26 | MYSIGRGATVSGSAITVTWTTVAGA |
| | Probability: | |
| 17135, | 0.848 AA1: | |
| 17136 | 27 AA2: 28 | MYSIRAKILSIVFLFIALTGVTFVIYS |
| | Probability: | |
| 42035, | 0.973 AA1: | |
| 42036 | 26 AA2: 27 | MYTCWKRRSRAASFSMCLRYSSSVVA |
| | Probability: | |
| 37591, | 1.000 AA1: | |
| 37592 | 21 AA2: 22 | MYWKQAFSFLVMLFATHSVHA |
| | Probability: | |
| 45475, | 0.803 AA1: | |
| 45476 | 20 AA2: 21 | MYYLFSLITGILISVMIAFN |
| 18319, | Probability: | |
| 18320 | 0.668 AA1: | MYYSYAVTLLILCPFIFLSSFIDA |

| SEQ ID NO: | Signalp Cleavage Site Probability: | Predicted Signal Sequence |
|---|---|---|
| | 24 AA2: 25 | |
| 78697 | 1.000 AA1: 30 AA2: 31 Probability: | LGAAAAGAAAGAGAPAAGAGAAAGALGAAA |
| 78715 | 1.000 AA1: 18 AA2: 19 Probability: | MKTNSRFLLAASLGLAMA |
| 78717 | 0.641 AA1: 37 AA2: 38 Probability: | MVFSWMPQLVKGARVTIALTVLAVSAGLVLSLFLALG |
| 78721 | 0.544 AA1: 37 AA2: 38 Probability: | MGSAWMLNCLCSLESAMNTSHTALFAILFAAMMPIVA |
| 78728 | 1.000 AA1: 20 AA2: 21 Probability: | MRKLILTATAVTAAIAGAFA |
| 78732 | 1.000 AA1: 23 AA2: 24 Probability: | MKNKLWVIVILTAFTLSATSCSA |
| 78763 | 1.000 AA1: 23 AA2: 24 Probability: | MVKTIFRTMLLALLSSAPLFAAA |
| 78787 | 0.679 AA1: 28 AA2: 29 Probability: | MICFTGVFMRRIAVIFIIACNITAGAFG |
| 78835 | 0.755 AA1: 21 AA2: 22 Probability: | MKKPRLLLAGAHSGAGKTSVA |
| 78865 | 0.740 AA1: 18 AA2: 19 Probability: | AKRNSGASGAGISANIQA |
| 78866 | 0.506 AA1: 34 AA2: 35 Probability: | LGLVFSKKLARFLKLFKHGFDFSILALGIALILA |
| 78870 | 1.000 AA1: 38 AA2: 39 Probability: | MAGVFFKSDGKFRLFSVFFSLILLVSMFSG CAGLAASA |
| 78887 | 0.954 AA1: 26 AA2: 27 Probability: | MGKDLKILKYLLCLLLVLPGLGRLEA |
| 78893 | 0.563 AA1: 25 AA2: 26 Probability: | MNQRALRADVLLLLTAGIWGFGFVA |
| 78907 | 0.996 AA1: 23 AA2: 24 | MSNKSVFKSAAAVILAALFLAIA |
| 78911 | Probability: | VRFPFALASMLAAANGYLSA |

| | | |
|---|---|---|
| | 0.999 AA1: | |
| | 20 AA2: 21 | |
| | Probability: | |
| | 0.861 AA1: | |
| 78925 | 22 AA2: 23 | MSILKEGYPFIIPAFIAGAAAA |
| | Probability: | |
| | 0.987 AA1: | |
| 78938 | 20 AA2: 21 | MRARFFAVVLISVWAFSACG |
| | Probability: | |
| | 0.571 AA1: | |
| 78958 | 31 AA2: 32 | AQIAAVRALSSGKDTVACSSTGNAASSLAGA |
| | Probability: | |
| | 0.918 AA1: | |
| 78996 | 22 AA2: 23 | MSIISCIGSGNMGFALMSAAKG |
| | Probability: | |
| | 0.632 AA1: | |
| 79003 | 20 AA2: 21 | MHSLKLILINASVFALFCFS |
| | Probability: | |
| | 0.573 AA1: | |
| 79030 | 22 AA2: 23 | VSVGGTKVLAPVAGVVIRYAVA |
| | Probability: | |
| | 0.999 AA1: | |
| 79035 | 25 AA2: 26 | MKNVLKLLGIIALVAVIGFSMSCGG |
| | Probability: | |
| | 0.967 AA1: | |
| 79036 | 22 AA2: 23 | MIKKAVLITVLANLLFCAEINA |
| | Probability: | |
| | 1.000 AA1: | |
| 79108 | 24 AA2: 25 | LKRLPITPFLLLMLLFCGTSFVYA |
| | Probability: | |
| | 0.596 AA1: | |
| 79135 | 30 AA2: 31 | VSTAVCLFWTAVLSFMPISELRGAIPFAVA |
| | Probability: | |
| | 0.820 AA1: | |
| 79164 | 23 AA2: 24 | SVSVVSAADTGGVSASSISAASA |
| | Probability: | |
| | 0.543 AA1: | |
| 79174 | 32 AA2: 33 | MYIKIPGILKKRPLYVQILFTVFAFLLMVVLS |
| | Probability: | |
| | 1.000 AA1: | |
| 79177 | 25 AA2: 26 | MSLYTFRKLCIAAIAALFIPLPSAA |
| | Probability: | |
| | 0.994 AA1: | |
| 79198 | 35 AA2: 36 | MNKRLSVLTNGIIKENPVLVLILGICPTLAVSTMA |
| | Probability: | |
| | 1.000 AA1: | |
| 79212 | 41 AA2: 42 | MTYLIKFFNKASIKRSITLLTRRYKILMLLLLIMTIGCSLV |
| | Probability: | |
| | 0.983 AA1: | |
| 79236 | 22 AA2: 23 | MKKVVFTAALIISVMLAVSCKS |
| 79240 | Probability: | GSPLRRLVKAAAKCAGGGGGGSPTRAQA |

| | | |
|---|---|---|
| | 0.822 AA1: 28 AA2: 29 Probability: 0.957 AA1: | |
| 79247 | 25 AA2: 26 Probability: 0.583 AA1: | YTLLRKPCLNKKPLSLCFLIACADG MDLQVLIAIGAGIAVFTGVGAGIGIGIATSGFLQAI |
| 79264 | 57 AA2: 58 Probability: 0.815 AA1: | SRQPEASGKMTPMFFVGVALA |
| 79285 | 29 AA2: 30 Probability: 0.996 AA1: | MKSPVSFKYLDVVASFFVVILVVSNIASS |
| 79288 | 28 AA2: 29 Probability: 0.975 AA1: | MKNKYTKLFTPLFALILFSSCMSINRNR |
| 79313 | 37 AA2: 38 Probability: 0.999 AA1: | LSFAIQCGRKMMKKKNGCVVFTLLLLISLIIPTKITA |
| 79319 | 29 AA2: 30 Probability: 1.000 AA1: | MGKTLKRCFAAFSITLFVISLAISCAGSA |
| 79356 | 22 AA2: 23 Probability: 0.773 AA1: | MKLNAKLLVLGAYLLMARPLFA |
| 79368 | 20 AA2: 21 Probability: 0.997 AA1: | MHDILVIGAGIAGLTVALKA |
| 79398 | 34 AA2: 35 Probability: 0.972 AA1: | MKIKLFLLLMIIAISAPMFGLDRVMDFGGLLSSA |
| 79408 | 23 AA2: 24 Probability: 0.994 AA1: | MAKKKIALVGAGQIGGTMALVLA |
| 79410 | 21 AA2: 22 Probability: 0.992 AA1: | MKKVLTIIVVIFVVFIGIVVA |
| 79431 | 23 AA2: 24 Probability: 1.000 AA1: | VPFVNDATALIRALAIALPLAFA |
| 79443 | 21 AA2: 22 Probability: 1.000 AA1: | MKRTVFLALALFALMAVGVYA |
| 79455 | 23 AA2: 24 Probability: 0.717 AA1: | MSIKRFLGFGFFLLLFAAAFVFG MADTKTGVQRKPILPLLGSSPHIASPVNSR |
| 79466 | 66 AA2: 67 Probability: 0.751 AA1: | MMMGNMLIAFFFVSVFSVINFGLPSLFNIIVAAAAA |
| 79471 | 27 AA2: 28 | MDEDNSKIRILIIAAVIIAIVALLIFA |
| 79476 | Probability: | MEAFFSVLPITIIVLIVSVFLTSLPSGTILMFLT |

| | | |
|---|---|---|
| | 0.844 AA1: 49 AA2: 50 Probability: | GAAMLIVGMGFFTLG |
| | 0.987 AA1: | |
| 79477 | 25 AA2: 26 Probability: | MKKLRFLGIIFLAAAISLFISCPME |
| | 0.586 AA1: | |
| 79499 | 29 AA2: 30 Probability: | MESGKKPGNKLRLADVLVVLFCLSGAAYS |
| | 0.507 AA1: | |
| 79502 | 22 AA2: 23 Probability: | LLTAGLGGMGGAQPLAATFAGA |
| | 0.991 AA1: | |
| 79510 | 25 AA2: 26 Probability: | MIKTILGFGWLIFAMIFLLPFALTA |
| | 0.999 AA1: | |
| 79514 | 20 AA2: 21 Probability: | MRIKFKLSIVVIAIMAVVVA |
| | 1.000 AA1: | |
| 79535 | 20 AA2: 21 Probability: | MKKFVCVFILLFFVSVLAFS |
| | 0.992 AA1: | |
| 79584 | 39 AA2: 40 Probability: | KKIVGKTLKNRKKNGFSDPVQAFLLVGALVAVSALISCA |
| | 0.995 AA1: | |
| 79594 | 22 AA2: 23 Probability: | MAKVKTNGLIIIVLWLAGIAAA |
| | 0.672 AA1: | |
| 79598 | 27 AA2: 28 Probability: | IASPIIVFHLAIMLIPAAALWIGCLMS |
| | 0.962 AA1: | |
| 79606 | 33 AA2: 34 Probability: | MNLPKLSVTRRVTVTMLILIMVLFGALSFVRLG |
| | 0.995 AA1: | |
| 79636 | 31 AA2: 32 Probability: | MKRFLIYLKSRPLGIASVIALIVLYLMMVFA |
| | 0.958 AA1: | VKTITKKKIKVKSSKKPTKKTKIIKTAPKVKKIVSAK |
| 79679 | 56 AA2: 57 Probability: | KTVKKPVIKVSAANNSAIK |
| | 0.969 AA1: | |
| 79727 | 16 AA2: 17 Probability: | GIGGLLITCIAGTVSA |
| | 0.995 AA1: | |
| 79745 | 23 AA2: 24 Probability: | MSKKQKIFLFSAGTALVALIVLG |
| | 0.533 AA1: | |
| 79747 | 15 AA2: 16 Probability: | MARGLAGSAPLTAIA |
| | 0.847 AA1: | |
| 79755 | 25 AA2: 26 | MPGLKKKHNRRFTLVVAGCMAQRLG |
| 79788 | Probability: | MKKYILALFILLFVVVFSAAQTA |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 23 AA2: 24 | |
| | Probability: | |
| | 0.957 AA1: | |
| 79804 | 25 AA2: 26 | MIRLRPGLYIPCGLCLALVLFLVFN |
| | Probability: | |
| | 0.995 AA1: | MIPLYKLEKLPRTQKLRKIAKIFTMAETRLSS |
| 79867 | 50 AA2: 51 | LLAGSLLASGLPASNLQV |
| | Probability: | |
| | 0.941 AA1: | |
| 79879 | 25 AA2: 26 | GVKTPDDPTSSLAVAAVGGAGGGAA |
| | Probability: | |
| | 0.505 AA1: | |
| 79918 | 26 AA2: 27 | LIRSAERRWPAAVGGVLVGFMTKEEA |
| | Probability: | |
| | 0.992 AA1: | |
| 79929 | 25 AA2: 26 | VNLKASITLFIILLIFSGSPVKIQA |
| | Probability: | |
| | 1.000 AA1: | |
| 79948 | 21 AA2: 22 | MKRTVFLSLALFALMAVGVYA |
| | Probability: | |
| | 1.000 AA1: | |
| 79964 | 36 AA2: 37 | MLTRKKDPVFIVKVCMLMLMGAFVFSFVSTAFNARA |
| | Probability: | |
| | 0.664 AA1: | |
| 79969 | 27 AA2: 28 | ELFLQKSYPPWQALLIVLGIVFGSVLS |
| | Probability: | |
| | 0.962 AA1: | |
| 80017 | 20 AA2: 21 | MSKFICTSAIDGAVAFVAQA |
| | Probability: | |
| | 1.000 AA1: | |
| 80046 | 37 AA2: 38 | MSQNKNQKRGSGIIYAASSFLLSLLPLLLMTTGNAVA |
| | Probability: | |
| | 0.995 AA1: | |
| 80048 | 17 AA2: 18 | MKRFFILLFFIPVLSFA |
| | Probability: | |
| | 0.956 AA1: | MQKVDLKISSMSLANLRAKPVRTVCLFVVV |
| 80073 | 44 AA2: 45 | AILAFTLFGGSILA |
| | Probability: | |
| | 0.973 AA1: | DSDVALWSRSEELLTPRGLYKLFKQCVKITV |
| 80097 | 51 AA2: 52 | MRFFVSVAVVFLLMSLPATA |
| | Probability: | |
| | 0.670 AA1: | |
| 80105 | 36 AA2: 37 | MDDYTIDKERRMKGNYRFSALIIAALIAIPAGYGQS |
| | Probability: | |
| | 0.751 AA1: | |
| 80113 | 36 AA2: 37 | VLFGTVWILFSKNILNDKWAKILMVACVSIDMLVFA |
| | Probability: | |
| | 0.975 AA1: | |
| 80118 | 27 AA2: 28 | MKPQMLFGAVMAIVGAFNNGAIGVALS |
| 80152 | Probability: | MLLFALSFASCASMASKASV |

| | | |
|---|---|---|
| | 0.991 AA1: 20 AA2: 21 Probability: 0.996 AA1: | |
| 80181 | 22 AA2: 23 | MKKTIFYLILISCIAVCGFSQT |
| | Probability: 0.535 AA1: | |
| 80192 | 26 AA2: 27 | MLKQISIKNLALIAGAEISFSNGFWA |
| | Probability: 0.830 AA1: | |
| 80198 | 44 AA2: 45 | MAAAFRKNKGRIDVREGFAIVTFSWLATTFIGAIPFFTFLSAHS |
| | Probability: 0.838 AA1: | |
| 80209 | 36 AA2: 37 | MSRFFKKQGFTLIELMVVIVIIGVLASLAIPRFTEA |
| | Probability: 0.585 AA1: | |
| 80234 | 21 AA2: 22 | MKYSNPANLAVVACPGGEIFA |
| | Probability: 0.996 AA1: | |
| 80245 | 25 AA2: 26 | MINQRKYITNLVLVALFASLISAGA |
| | Probability: 0.971 AA1: | |
| 80257 | 29 AA2: 30 | MKKLPEKALFIIFFALAFSQFPKNALLHA |
| | Probability: 0.993 AA1: | |
| 80265 | 28 AA2: 29 | SRSSFFLLNLALLVLSCLMDIFSAILVG |
| | Probability: 0.625 AA1: | |
| 80269 | 41 AA2: 42 | MAEVLCIALTFTKFFRARPIILLMALVAYLNFGTAIVTATA |
| | Probability: 1.000 AA1: | |
| 80315 | 19 AA2: 20 | MKKIIALLIGLSVTAICFA |
| | Probability: 0.991 AA1: | |
| 80343 | 20 AA2: 21 | GYTKGVAAVAGGVSGFAGVA |
| | Probability: 0.993 AA1: | |
| 80360 | 36 AA2: 37 | MSERRKALAAIIACVLFWGFSFISIKITVAVFPPMS |
| | Probability: 1.000 AA1: | MQIFQKSNLNERRTSIKRKFIVFSSVLFLAIVLLGS |
| 80370 | 52 AA2: 53 | GAFLLSMRQIVRNSTA |
| | Probability: 1.000 AA1: | |
| 80381 | 20 AA2: 21 | MIKKIYYLLILLIISAPALA |
| | Probability: 0.995 AA1: | |
| 80382 | 32 AA2: 33 | MKLSVKISLLVGIVLLASTVALLFVVGLKVSG |
| | Probability: 0.988 AA1: | |
| 80387 | 39 AA2: 40 | MNKKRLFDWWRRGSAFVWLSGSTLALSLLMVAGLVLLLA |
| 80399 | Probability: | MKKLILFLFIISSNLFG |

| | | |
|---|---|---|
| | 0.996 AA1: 17 AA2: 18 Probability: 0.910 AA1: | |
| 80410 | 21 AA2: 22 Probability: 0.999 AA1: | MKTNKSGWALVLSGGGAMGLA |
| 80425 | 23 AA2: 24 Probability: 1.000 AA1: | MGKFARFGLTLLLTTATIVGAVA |
| 80429 | 28 AA2: 29 Probability: 0.981 AA1: | LFVGVTLKNAFKSAALLAATLLAAQSFA |
| 80459 | 22 AA2: 23 Probability: 0.999 AA1: | MNKKTVLCILILIMVYASCSQK |
| 80460 | 24 AA2: 25 Probability: 0.819 AA1: | MDQIGTRKLSVVLLVLLVAGSLFA |
| 80471 | 29 AA2: 30 Probability: 0.994 AA1: | MKNNFADKVIIQALLCATLLLVLVTGCDS |
| 80472 | 37 AA2: 38 Probability: 1.000 AA1: | MGIHSNSAISTWLVRFLAALTVILAGLIGLGLGLSLA |
| 80496 | 32 AA2: 33 Probability: 0.992 AA1: | MRNYEKNIFTTRLIFAGFGVLLLFLFACTAQK LRILIKYLKPFTLLLFLAVMILFLQVFCDLSLPNL |
| 80501 | 49 AA2: 50 Probability: 0.672 AA1: | MSRLVNSILAQSAA |
| 80502 | 25 AA2: 26 Probability: 0.552 AA1: | MRKLYNYPWLIVAVIAVITVFFSFQ |
| 80515 | 19 AA2: 20 Probability: 0.989 AA1: | AKTGLASVAFPNISTGVYG |
| 80526 | 24 AA2: 25 Probability: 1.000 AA1: | MKKGIAFLTLVVFTLFFITNQVFS |
| 80550 | 20 AA2: 21 Probability: 0.887 AA1: | MKNKIIFFTLLAGAALALYA |
| 80560 | 21 AA2: 22 Probability: 0.999 AA1: | YQESDRLAAVSAVLRALGADA |
| 80565 | 24 AA2: 25 Probability: 0.724 AA1: | MRGILRFASLSMALWIFLPFALFA |
| 80576 | 21 AA2: 22 | MKELQYDAVVIGAGAAGMAAA |
| 80578 | Probability: | MVKKISGIALFLISFPMFWVSACA |

| | | |
|---|---|---|
| | 0.998 AA1: 24 AA2: 25 Probability: | |
| | 0.963 AA1: | |
| 80583 | 34 AA2: 35 | MDKNRVPKTRPFAGKNNINSLACIALFPCLPVFS |
| | Probability: | |
| | 0.831 AA1: | |
| 80597 | 17 AA2: 18 | MNGLLLLLISVITLSLA |
| | Probability: | |
| | 0.754 AA1: | |
| 80622 | 33 AA2: 34 | GAPLTDSAENWAIRFIQAVVAVILMAVFLYVFA |
| | Probability: | |
| | 0.993 AA1: | MSERRKALAAIIACVLFWGFSFISIKITVAVF |
| 80672 | 45 AA2: 46 | PPMSLGMLRFAIA |
| | Probability: | |
| | 0.577 AA1: | DPFAAYICAGLVGVYSFHFLINVGMTMGIMPITGIP |
| 80680 | 50 AA2: 51 | LLFMSYGGSSLLSA |
| | Probability: | |
| | 1.000 AA1: | |
| 80689 | 27 AA2: 28 | MSKKKAACLLALPPILAAAFSFILPYA |
| | Probability: | |
| | 0.982 AA1: | |
| 80715 | 22 AA2: 23 | MKISLWKAMCTGLLCAAVLLTA |
| | Probability: | |
| | 0.530 AA1: | |
| 80739 | 19 AA2: 20 | MKKRIAVLGATGSIGKSSL |
| | Probability: | |
| | 0.831 AA1: | |
| 80745 | 43 AA2: 44 | MVVQYAIFWHITLKSKSGGMMTLFALIGLLPMCVIAPFAGVWA |
| | Probability: | |
| | 0.660 AA1: | |
| 80768 | 37 AA2: 38 | MYVVEHALLNLGKNKSRNILLGVIIFAIITATVVALA |
| | Probability: | |
| | 0.969 AA1: | |
| 80780 | 29 AA2: 30 | CRPSLLGGSDVFSTASAAPSASTGSAAGA |
| | Probability: | |
| | 0.624 AA1: | |
| 80806 | 18 AA2: 19 | MKFTRVDITVPPTLVAFA |
| | Probability: | |
| | 0.579 AA1: | LIANWVSLNLARFSYAALLTRALLGFVGALPWLLA |
| 80807 | 45 AA2: 46 | AFGFNSPAVR |
| | Probability: | |
| | 0.997 AA1: | |
| 80815 | 23 AA2: 24 | LNIILLLSPGFLLFQSCATGSFA |
| | Probability: | |
| | 0.778 AA1: | |
| 80832 | 13 AA2: 14 | MAIFLTALGAALA |
| | Probability: | |
| | 0.736 AA1: | |
| 80839 | 20 AA2: 21 | MRLRLASTVTANSALVLKMA |
| 80849 | Probability: | MRTGSAADLMLGLSVFLRRG |

| | | |
|---|---|---|
| | 0.692 AA1: | GYRGGGRRCFALPAV |
| | 47 AA2: 48 | ALLFAAALPAAA |
| | Probability: | |
| | 1.000 AA1: | MLNNKKDPVFIVKVCMLMLVGAFVFS |
| 80851 | 36 AA2: 37 | SVSAAFNARA |
| | Probability: | |
| | 0.790 AA1: | |
| 80852 | 15 AA2: 16 | MAGFLITCIAGTVSA |
| | Probability: | |
| | 0.701 AA1: | |
| 80853 | 29 AA2: 30 | MALNCGIVGLPNVGKSTIFSALSSAPAEA |
| | Probability: | |
| | 0.784 AA1: | |
| 80868 | 20 AA2: 21 | MKILLPILTIFTCLGFAGHG |
| | Probability: | |
| | 1.000 AA1: | |
| 80893 | 21 AA2: 22 | MKKLPLLFIATLFAAIPAAHA |
| | Probability: | |
| | 0.982 AA1: | |
| 80901 | 23 AA2: 24 | LIIVTNIAGSLLLLLAISASLWL |
| | Probability: | |
| | 1.000 AA1: | |
| 80909 | 33 AA2: 34 | MKKTKWVQVVFKRRVLVILLLSVQAALIALAIA |
| | Probability: | |
| | 0.997 AA1: | |
| 80914 | 37 AA2: 38 | MNFKWIWFVAARYIFRKRKKSPAPVLSILGIAVGVLA |
| | Probability: | |
| | 0.549 AA1: | |
| 80926 | 18 AA2: 19 | VRNVLFVGATATGSGSLG |
| | Probability: | |
| | 0.705 AA1: | |
| 80975 | 31 AA2: 32 | VAALNSGAVALIGGGTGSPFFTTDTAAAVRS |
| | Probability: | |
| | 0.632 AA1: | |
| 80984 | 22 AA2: 23 | TAATAAYASDAAYAASFTAAYA |
| | Probability: | |
| | 0.981 AA1: | |
| 80986 | 19 AA2: 20 | MSRLAAASIIALFFLASCS |
| | Probability: | |
| | 0.699 AA1: | |
| 81007 | 23 AA2: 24 | MRLIHIYLFVSVLVLTLIAIPRA |
| | Probability: | |
| | 0.951 AA1: | LLPLFDLLRLGLFLFRRNVLTPRFTAAGIGAMLCLC |
| 81030 | 47 AA2: 48 | AVIIVCGALHA |
| | Probability: | |
| | 0.999 AA1: | |
| 81049 | 20 AA2: 21 | MKKGIIAVILILGAVFCAFG |
| | Probability: | |
| | 0.612 AA1: | |
| 81123 | 26 AA2: 27 | MQHVLLSVNTLVVSAIFLIIAASIDA |
| 81125 | Probability: | MKKNICLATIALILVVSCSS |

| | | |
|---|---|---|
| | 0.997 AA1: 20 AA2: 21 Probability: 0.996 AA1: | |
| 81135 | 34 AA2: 35 Probability: 0.997 AA1: | MTYKIIFRTVSFSATAAFLMCAGCADFGTGGGSA |
| 81159 | 33 AA2: 34 Probability: 0.937 AA1: | VNRKNAPDYRPFLRLLIVILLVFLGFSMLPAAE |
| 81206 | 26 AA2: 27 Probability: 0.508 AA1: | VRDNVTVTATAASCYTFIGWSGAATG |
| 81242 | 26 AA2: 27 Probability: 0.629 AA1: | MTGRPAGSLGAPALTFWTTGLSYRMA |
| 81254 | 18 AA2: 19 Probability: 0.603 AA1: | MTDNNIGPVLLGAVIANA |
| 81310 | 21 AA2: 22 Probability: 0.671 AA1: | MIRVMVFILISVLFLAFIIVN |
| 81354 | 26 AA2: 27 Probability: 0.677 AA1: | MSSNTIFIVLIPIAAIAAFIVGFFWR |
| 81370 | 41 AA2: 42 Probability: 0.822 AA1: | MSDRVIHTSDTKTLIKNNIRNYSMFIMLALIMIIFAPLTKG |
| 81424 | 22 AA2: 23 Probability: 1.000 AA1: | DRGSWFQPVTALWIFLPFALFA |
| 81426 | 30 AA2: 31 Probability: 1.000 AA1: | VNNFWKTVSGSALAALALVCVFGIAAAVSA |
| 81448 | 21 AA2: 22 Probability: 0.965 AA1: | MKKYILFAVLLLLLLADAACA |
| 81485 | 19 AA2: 20 Probability: 0.980 AA1: | MMVMLFLLGLAVFGLMALA |
| 81494 | 29 AA2: 30 Probability: 0.773 AA1: | MEKTVKVIMIAIVAAVIVCVTVIGLSLAA |
| 81510 | 20 AA2: 21 Probability: 0.838 AA1: | MHDILVIGAGIAGLTVALKA<br>MDAIKPGSGTNFKDWVYTDKFRFYKITFFTEKLVM |
| 81532 | 55 AA2: 56 Probability: 0.581 AA1: | RKKLLLSIAIGLTLIAQAFA |
| 81577 | 27 AA2: 28 | VKANKTKMRIRSTTVLAVRRNGKVAMA |
| 81593 | Probability: | MVNGTSRYSGSFSISGTRITFSISIGSASATFSLS |

| | | |
|---|---|---|
| | 0.827 AA1: 35 AA2: 36 Probability: | |
| | 0.902 AA1: | |
| 81611 | 37 AA2: 38 | MVIMLFLGFKVNRRTGHALIAAIGFAALSLTGLSLTG |
| | Probability: 1.000 AA1: | |
| 81615 | 23 AA2: 24 | MKRKYIRCILFMVFLFSLLPLYA |
| | Probability: 0.966 AA1: | |
| 81653 | 27 AA2: 28 | LIIAGIAGMPIFAAIGGIALIMLQAAG |
| | Probability: 0.983 AA1: | |
| 81661 | 30 AA2: 31 | MKNTLNVLLVRVLGIIALVAVIGFSMTACG |
| | Probability: 0.646 AA1: | GFSPAYREFVISHPTVTTWPTAATITYGAALSTSA |
| 81668 | 42 AA2: 43 | LSGGAGA |
| | Probability: 0.977 AA1: | |
| 81671 | 24 AA2: 25 | MTQIQIINSAIGMVSSFLIAMALG |
| | Probability: 0.997 AA1: | |
| 81700 | 37 AA2: 38 | MNYKWIWFIAARYIFRKRKISPAPVLSILGIATGVLA |
| | Probability: 1.000 AA1: | |
| 81717 | 23 AA2: 24 | MIKKIIHGFFFVLTLSFSLVFSA |
| | Probability: 0.785 AA1: | |
| 81744 | 25 AA2: 26 | MKYDVILAGVGGQGVLSVASIIAQA |
| | Probability: 1.000 AA1: | |
| 81784 | 22 AA2: 23 | MKTIMKICTVCLLLASSTTLFA |
| | Probability: 0.999 AA1: | |
| 81795 | 18 AA2: 19 | MKKIVVLFFLIAVGGVYA |
| | Probability: 0.991 AA1: | |
| 81815 | 18 AA2: 19 | MISRVAYLIILTAFFAAA |
| | Probability: 0.995 AA1: | |
| 81824 | 22 AA2: 23 | MYGIKRALAPPALRAFVVTAAA |
| | Probability: 0.728 AA1: | |
| 81847 | 30 AA2: 31 | VREVKISENYLERAVLIYLFIFCFPLCAQA |
| | Probability: 0.988 AA1: | |
| 81868 | 25 AA2: 26 | MGKKNITYLAVLAALLTPAIVSCNS |
| | Probability: 0.977 AA1: | |
| 81877 | 20 AA2: 21 | MSTGVIVFAIFLICFFALFA |
| 81888 | Probability: | MKNIKDKLILITGGASGIGQLMA |

| | | |
|---|---|---|
| | 0.977 AA1: 23 AA2: 24 Probability: 0.966 AA1: | |
| 81891 | 38 AA2: 39 Probability: 0.945 AA1: | MFDRKVDPIWVLKGCMILLVAAFLFNFIGAAFSPRPAA |
| 81894 | 24 AA2: 25 Probability: 0.898 AA1: | MKLYSRGQVLFFSFLSAVIVLLFA |
| 81920 | 21 AA2: 22 Probability: 0.816 AA1: | VKGLSLVLHSAVTIMTPILLA |
| 81921 | 19 AA2: 20 Probability: 0.843 AA1: | MLIGLGGGAASSMATGSGG |
| 81923 | 25 AA2: 26 Probability: 0.828 AA1: | MKVWLKLLIGSILGVIIGLHLPADN |
| 81969 | 18 AA2: 19 Probability: 0.979 AA1: | LQKQTITTFSASPKTASA MRIKLISTYKIIRLFLQAKLALPVRLALFALLALPALCC |
| 81986 | 49 AA2: 50 Probability: 1.000 AA1: | LFVTCKDSGG |
| 82049 | 20 AA2: 21 Probability: 0.616 AA1: | MKKIFFVVCLVFFIAGFAVA |
| 82050 | 28 AA2: 29 Probability: 0.963 AA1: | MGRTPGIICVCSFLAAVIVTLTNLNLSS |
| 82125 | 17 AA2: 18 Probability: 0.943 AA1: | MTIREALAGGSAALAAA |
| 82140 | 25 AA2: 26 Probability: 0.996 AA1: | LIIAIVAVIVTGVAALLLRQASSIA |
| 82181 | 23 AA2: 24 Probability: 1.000 AA1: | MANRKLLTVMLAVALVSAMTACK |
| 82187 | 20 AA2: 21 Probability: 0.696 AA1: | MKKIITTVIAVLIYAGAVFA |
| 82222 | 34 AA2: 35 Probability: 0.511 AA1: | MQKKTIKLGTVKTANFRNNAFLCIGTGRMGLALQ |
| 82224 | 19 AA2: 20 Probability: 0.502 AA1: | MFSTSSLVTAFPAAQTGMG |
| 82234 | 19 AA2: 20 | LGKLGLSMSGAINVFFRQA |
| 82256 | Probability: | VNSLIYIYAVCMKRRIKILDIIIIFIAVSLTISSAYA |

| | | |
|---|---|---|
| | 0.988 AA1: 37 AA2: 38 Probability: 0.910 AA1: | |
| 82271 | 33 AA2: 34 | MSMVKTVVSRPTTFLIIFLLLIGLGFFAMFNLS |
| | Probability: 0.959 AA1: | |
| 82284 | 24 AA2: 25 | MAFVAKRLVITLITLFLVSVITFA |
| | Probability: 0.852 AA1: | |
| 82324 | 24 AA2: 25 | MKIPGRHQALNAAAAVALTASLLK |
| | Probability: 0.629 AA1: | |
| 82350 | 38 AA2: 39 | MGPDMTKKIKVVKRFLKQAKFTNLVLYAALVLISYQFM |
| | Probability: 0.725 AA1: | |
| 82375 | 36 AA2: 37 | MSTEYQKIVNALKRRRKGAVAADICAATALPLSSVN |
| | Probability: 0.842 AA1: | |
| 82393 | 23 AA2: 24 | LNTERKVLVTGASGGIGRAIALA |
| | Probability: 0.912 AA1: | |
| 82407 | 19 AA2: 20 | TERMLMFGAVMAITQAFNV |
| | Probability: 0.800 AA1: | |
| 82442 | 24 AA2: 25 | MGNGQRVLFLLIIIPLFFILTARQ |
| | Probability: 0.945 AA1: | |
| 82467 | 18 AA2: 19 | MTRKQASTISIMILAVVA |
| | Probability: 0.952 AA1: | |
| 82469 | 18 AA2: 19 | MIIATVIICAFFTSFTGA |
| | Probability: 0.995 AA1: | |
| 82481 | 30 AA2: 31 | MNSLPKRRFLFAVSALFAAALYILIRYAMR |
| | Probability: 1.000 AA1: | |
| 82499 | 30 AA2: 31 | MEKIIFVSKLRVSPWLIILLFIAGIASAYA |
| | Probability: 0.999 AA1: | |
| 82503 | 25 AA2: 26 | MAIAGASRFPVFILSALLWAAGASA |
| | Probability: 0.951 AA1: | |
| 82515 | 31 AA2: 32 | MKGIKPVLLAISTALVLKLFFFDFIIAQGHS |
| | Probability: 1.000 AA1: | |
| 82527 | 23 AA2: 24 | YRREIIRAAAFALLASISLTAQS |
| | Probability: 0.541 AA1: | |
| 82554 | 31 AA2: 32 | MGPNITSTQTRLLRATWLPTILLLSGRTRIA |
| 82572 | Probability: | MPAILNLIWLLPILAVFA |

| | | |
|---|---|---|
| | 0.678 AA1: 18 AA2: 19 Probability: 0.881 AA1: | |
| 82631 | 25 AA2: 26 Probability: 0.545 AA1: | MYSIGRGATVSGSAITVTWTTVAGA |
| 82653 | 17 AA2: 18 Probability: 0.508 AA1: | MISLVILAALYVVMTFA |
| 82654 | 16 AA2: 17 Probability: 0.783 AA1: | MRRVVITGSSVISALG |
| 82701 | 26 AA2: 27 Probability: 0.714 AA1: | MANKRSSVTKTFGILTAGGDCPGLNA |
| 82706 | 25 AA2: 26 Probability: 0.858 AA1: | VGTWAGALLGTCTVTSYVESTAGVA |
| 82717 | 23 AA2: 24 Probability: 0.874 AA1: | MFINPPLFNLCLTISAMAGFIVS |
| 82762 | 38 AA2: 39 Probability: 0.540 AA1: | MTRTKAGYVEGAVSIIINLLLFILKMWASVISGSIALA |
| 82774 | 26 AA2: 27 Probability: 0.750 AA1: | MFYTRWVLTLTVCLPKITLFKAGTHA |
| 82803 | 18 AA2: 19 Probability: 1.000 AA1: | MKLAGLTVLTAALLLSCS |
| 82806 | 32 AA2: 33 Probability: 0.538 AA1: | MNNRNRIIAGVIISLLIIIAAASLGSSGISLR |
| 82870 | 34 AA2: 35 Probability: 0.999 AA1: | MHNAKFIVPNAFTSINFLLGVASICCSASLFEIK |
| 82887 | 21 AA2: 22 Probability: 0.996 AA1: | MQKAAALFLILFFGVFVLLFA |
| 82925 | 33 AA2: 34 Probability: 0.689 AA1: | MNLFKNASISKKLIMLLVVFILGYSIFGLVAFS |
| 82936 | 27 AA2: 28 Probability: 0.970 AA1: | MIRSAEDCLTKLWALILASVLASMGLA |
| 82950 | 23 AA2: 24 Probability: 0.992 AA1: | VNKQSLAKRVLQTAPLMILSVFA |
| 82952 | 26 AA2: 27 | MKIRLTLIIIASLLLFQGASPWEGSA |
| 82958 | Probability: | MKIKIKLSIVVIAIMAVVVA |

| | | |
|---|---|---|
| | 0.999 AA1: 20 AA2: 21 Probability: | |
| | 0.658 AA1: | |
| 82967 | 25 AA2: 26 | AEGRYFEAILCGGMIFMAMVTFVLA |
| | Probability: 0.990 AA1: | |
| 82978 | 23 AA2: 24 | MSDRFKSIYFLFLPAGIVLLALA |
| | Probability: 0.991 AA1: | MRHIPKAAFLIPVLIKLFRLLTHSHKIYLIILFVM |
| 82987 | 57 AA2: 58 | TVGFALVETLGMSVIMPFISLA |
| | Probability: 0.815 AA1: | |
| 83011 | 24 AA2: 25 | MGIVLIILAFVLLIIGIIGAVVPA |
| | Probability: 1.000 AA1: | |
| 83016 | 20 AA2: 21 | MRFVFSAVILILLGSFPAAA |
| | Probability: 0.778 AA1: | |
| 83021 | 23 AA2: 24 | WFPKTDWPRVTLLLSGMSTQAQL |
| | Probability: 0.988 AA1: | |
| 83027 | 25 AA2: 26 | MGNPPFTQIRVAGLMFAGLNGGSLA |
| | Probability: 1.000 AA1: | |
| 83031 | 20 AA2: 21 | MRKLILTTAAVAAAIAGAFA |
| | Probability: 0.955 AA1: | |
| 83043 | 20 AA2: 21 | MWVIYALLSAVFAALTSILA |
| | Probability: 0.999 AA1: | |
| 83046 | 29 AA2: 30 | MKKRIMLAAVMTFFVTLIACKHIDSLAQA |
| | Probability: 0.999 AA1: | |
| 83055 | 39 AA2: 40 | MKRIFRWIFAGIVAIPVMAALFGAVVMLLWNALMPQLFA |
| | Probability: 0.625 AA1: | |
| 83067 | 36 AA2: 37 | MGKNQQDKVRFSIGNKLIIIISIIVILSLGSITALV |
| | Probability: 1.000 AA1: | |
| 83084 | 34 AA2: 35 | MRHKSKSLFSASESLLFAAIALILSFAHIAPVTA |
| | Probability: 0.957 AA1: | |
| 83091 | 27 AA2: 28 | MAILSTVLLVFFVIIAVLLVLLVLVQN |
| | Probability: 0.994 AA1: | |
| 83150 | 24 AA2: 25 | MIRKPYKLFLFSLLLSLALSYLPA |
| | Probability: 0.843 AA1: | |
| 83172 | 39 AA2: 40 | VVLRAVFVIAPIAWIYARADAGSISATLASASVPAIACA |
| 83177 | Probability: | MKLSALAFLGIIGLFIVTPLFS |

| | | |
|---|---|---|
| | 0.985 AA1: 22 AA2: 23 Probability: 0.999 AA1: | |
| 83180 | 31 AA2: 32 Probability: 0.904 AA1: | MKTILKKNYIFVAAAILIVSAYFISGVLVFA |
| 83203 | 20 AA2: 21 Probability: 0.985 AA1: | MGIAKFLVSLAVFAMFLACG |
| 83223 | 27 AA2: 28 Probability: 0.939 AA1: | ILGNGASFGSVASFGSVASSGSGASFG |
| 83242 | 25 AA2: 26 Probability: 0.726 AA1: | VIKILKYLKPKEWLMAALSLIFIVA |
| 83252 | 25 AA2: 26 Probability: 0.999 AA1: | LSIASTSIRRPVLTIVLSALIVLFG |
| 83253 | 20 AA2: 21 Probability: 0.813 AA1: | MRKGLSFCIIFLAAFVFIYA |
| 83282 | 41 AA2: 42 Probability: 0.693 AA1: | LFKTLVPPLFKIIPLPLITILILLFGSLVGKKFFAVQSAYS |
| 83283 | 39 AA2: 40 Probability: 0.997 AA1: | MSKKLSQSIAFWIFRILGVLVVGILLWILGFIIINGISA |
| 83292 | 26 AA2: 27 Probability: 1.000 AA1: | MTKFTILLAIVSIIIAAVIIFFTLQS |
| 83293 | 39 AA2: 40 Probability: 0.999 AA1: | MRGFRHSSGMNKTVRTVVWSVMAAVLAVAGSGWLFGASA |
| 83299 | 29 AA2: 30 Probability: 1.000 AA1: | MRFGRKKAGKLFPALALFLFSLSFPNAVK |
| 83302 | 22 AA2: 23 Probability: 0.886 AA1: | MRRRIIFSALLLLALVPQAAYT |
| 83306 | 28 AA2: 29 Probability: 1.000 AA1: | PPSVRFGRGAVVYTIALSPGFPPPPAFA |
| 83322 | 19 AA2: 20 Probability: 0.869 AA1: | VFIRFALCAFLLLPIFALA |
| 83325 | 21 AA2: 22 Probability: 1.000 AA1: | ARTGGITSLLGLCYLSVAGFS |
| 83329 | 30 AA2: 31 | LKPSPVRKFAAFLPFAVLILISAATVKTTA |
| 83355 | Probability: | LVRNTITSLWPIKQGEAARHARRATWRR |

| | | |
|---|---|---|
| | 0.637 AA1: | FSFSQSVFAITIV |
| | 59 AA2: 60 | FLFLPLLVLILYSFNDSA |
| | Probability: | |
| | 0.694 AA1: | |
| 83366 | 27 AA2: 28 | MTYFIHAIMIFLGLLAVVPIYFLLLNA |
| | Probability: | |
| | 1.000 AA1: | MKLSQMSIKSKMLVFSGLFFMTVAVLGTA |
| 83392 | 44 AA2: 45 | AFFFSMRQIVRSSAA |
| | Probability: | |
| | 1.000 AA1: | |
| 83404 | 31 AA2: 32 | MKTANKIPANIWGLLTIIVILGVMMPMKATA |
| | Probability: | |
| | 0.998 AA1: | |
| 83417 | 20 AA2: 21 | MKYKLLFFVIFISFVFSCYS |
| | Probability: | |
| | 0.928 AA1: | |
| 83420 | 23 AA2: 24 | MNHPGISLILATGGAGMVKSAYS |
| | Probability: | |
| | 0.999 AA1: | |
| 83421 | 37 AA2: 38 | MNLKWIWFIASRYIFRKRKKSPAPVLSILGIAVGVLA |
| | Probability: | |
| | 0.996 AA1: | |
| 83424 | 34 AA2: 35 | MKRAMPENKGRRKPYHLSFVIFYLLFFICSAAWA |
| | Probability: | |
| | 0.586 AA1: | |
| 83430 | 19 AA2: 20 | MAACIGASIAGLKSMSATS |
| | Probability: | |
| | 0.798 AA1: | |
| 83440 | 31 AA2: 32 | MRKFYNYPWLIVAAIAIVTVFFGFQLPRAQL |
| | Probability: | |
| | 1.000 AA1: | |
| 83450 | 21 AA2: 22 | MKRKVILALALFALMAVGVYA |
| | Probability: | |
| | 0.998 AA1: | |
| 83456 | 19 AA2: 20 | MKTAIFILLCIICINGVWA |
| | Probability: | |
| | 0.912 AA1: | |
| 83473 | 26 AA2: 27 | LIESAIINEILAAAPAFLLIAVRALA |
| | Probability: | |
| | 0.903 AA1: | |
| 83487 | 19 AA2: 20 | MIYTVLIFVGAAIAGALVS |
| | Probability: | |
| | 0.993 AA1: | MDAQKLRGNRSFERRSKGLRFLTKLKAYPL |
| 83534 | 45 AA2: 46 | LVVPLVLGAMRKAQA |
| | Probability: | |
| | 0.988 AA1: | MRIGGRNPAFFVHTKAGGRLPLPQFLFMLL |
| 83560 | 40 AA2: 41 | VLMFVPSAFA |
| | Probability: | |
| | 0.994 AA1: | |
| 83597 | 21 AA2: 22 | MKRFNLILLIVLFGFLFSCAG |
| 83601 | Probability: | VVPFSFGLFFLVAWFLWA |

| | | |
|---|---|---|
| | 0.512 AA1: 18 AA2: 19 Probability: 1.000 AA1: | |
| 83612 | 31 AA2: 32 Probability: 0.754 AA1: | LQSSSISKFTALLLYVASILTGAAAVRPASA |
| 83613 | 23 AA2: 24 Probability: 0.513 AA1: | MNRKSAKKNLFITAPGGYGKTTA |
| 83656 | 18 AA2: 19 Probability: 0.999 AA1: | MTAQTAVIGAGAWGTAVA GIIAASVPYTFLFLSIIYYKYRKGGFSLKSSV |
| 83668 | 50 AA2: 51 Probability: 0.937 AA1: | LVLIFLFTFFLFCAPVQA |
| 83677 | 24 AA2: 25 Probability: 0.863 AA1: | VQKGIKSSSVSSSAAAGHGASDLA |
| 83682 | 21 AA2: 22 Probability: 0.998 AA1: | VVTGAPSGSIVMYSTSANSVA |
| 83684 | 37 AA2: 38 Probability: 0.525 AA1: | MEKNKTQERETSIKGRFIAFSAVMFLVIFIGGSAAFA |
| 83694 | 17 AA2: 18 Probability: 1.000 AA1: | VIVFLTVLTFTASRAES |
| 83704 | 25 AA2: 26 Probability: 0.998 AA1: | MSNRNRIVAGAVISLLIIIAAASLG |
| 83711 | 19 AA2: 20 Probability: 0.854 AA1: | MKKRVLIFLLLNVSCIIYA |
| 83729 | 20 AA2: 21 Probability: 0.544 AA1: | MANPLFQRLSLLIGAGAADA |
| 83752 | 17 AA2: 18 Probability: 0.681 AA1: | MNGLFLLVISVVVLSLA |
| 83761 | 25 AA2: 26 Probability: 0.617 AA1: | MGLFDRNLLALSAHNPALCSQLSLA |
| 83777 | 16 AA2: 17 Probability: 0.970 AA1: | MTPSGTLYSATIPAQA |
| 83809 | 37 AA2: 38 Probability: 0.531 AA1: | MRYRLRSKLALFTIALVLLVDLLISAPLFFMIIKTQQ |
| 83827 | 37 AA2: 38 | MKKRNSKNRMVTPQANGSTMKEPLPPALLSAPGQRPA |
| 83835 | Probability: | MRQKLCLLACLVLLCGCSRNG |

| | | |
|---|---|---|
| | 0.998 AA1: 21 AA2: 22 Probability: | |
| | 0.940 AA1: | |
| 83899 | 26 AA2: 27 | MKKKNECVVFTLLLLISLIIPIKITA |
| | Probability: 0.818 AA1: | |
| 83907 | 25 AA2: 26 | MINLNVLRKFLLPLVIIAASLCVLT |
| | Probability: 0.632 AA1: | |
| 83923 | 36 AA2: 37 | MSQNPTEVYWNSDRKMRFLFAASGLAAFFALLYYLS |
| | Probability: 0.976 AA1: | |
| 83941 | 27 AA2: 28 | MAESFGRWLPLGVLGASLLFGFSSSLA |
| | Probability: 0.715 AA1: | |
| 83949 | 30 AA2: 31 | KSGTNRPSIVGSYDLFIAGGGCAGTGAAIA |
| | Probability: 1.000 AA1: | |
| 83955 | 21 AA2: 22 | MLTLKLTIALACTLLCTTIFA |
| | Probability: 1.000 AA1: | |
| 83975 | 21 AA2: 22 | VPRPIAFSFALLVIFGAMALA |
| | Probability: 0.549 AA1: | |
| 83982 | 27 AA2: 28 | CEYWTAIPAGTAAGASTFGTSGIFAIA |
| | Probability: 0.873 AA1: | |
| 84068 | 30 AA2: 31 | MNIETSLFYDLKILLPVILVSLGAIFSLVA |
| | Probability: 0.977 AA1: | |
| 84069 | 23 AA2: 24 | MKKPFAYSATFICTIIGAGFASG |
| | Probability: 0.635 AA1: | |
| 84127 | 30 AA2: 31 | VISNGYSFIESLVYALGVALGFMLSLLLLA |
| | Probability: 0.983 AA1: | |
| 84128 | 25 AA2: 26 | FIERAELGLILAVFAAAACVMLAFV |
| | Probability: 0.639 AA1: | |
| 84144 | 22 AA2: 23 | MNKPETIRSAAKFAASSLVRHA |
| | Probability: 0.738 AA1: | |
| 84177 | 31 AA2: 32 | MENKKYLKFKEKYGPYALVAGGSNGLGAAYA |
| | Probability: 0.785 AA1: | |
| 84210 | 25 AA2: 26 | MAPVYEMLTITVIISVVIMVLVVLA |
| | Probability: 0.946 AA1: | |
| 84214 | 18 AA2: 19 | MRWIVCVLSASLESSAWA |
| 84218 | Probability: | CQIRTRGFPDLVPVLAVAAASA |

| | | |
|---|---|---|
| | 0.929 AA1: 22 AA2: 23 Probability: 1.000 AA1: | |
| 84246 | 22 AA2: 23 Probability: 0.958 AA1: | VKIKLLGFLFLLSFPLFSLSFN |
| 84256 | 27 AA2: 28 Probability: 0.996 AA1: | MAILSTVLLVFFVIIAILLVLLVLVQN |
| 84258 | 21 AA2: 22 Probability: 0.996 AA1: | MRTVNLLTVIATIGTFCLSNA |
| 84354 | 25 AA2: 26 Probability: 0.798 AA1: | MKSILKRSYVFVAGAALIVSAYFIA |
| 84365 | 26 AA2: 27 Probability: 0.982 AA1: | MYTFNAYAVAVVALVILVFTTMMFLA |
| 84373 | 26 AA2: 27 Probability: 0.999 AA1: | MRPLIHALTKFFAALLVIALPFVSCS |
| 84418 | 40 AA2: 41 Probability: 0.999 AA1: | MEKNNKRRAIDNIKFSIGKKLIMIISIMVIISLGSITALA |
| 84462 | 30 AA2: 31 Probability: 0.999 AA1: | MKALVAWLKRKKSAFCILAGIISVSLFVIS |
| 84486 | 20 AA2: 21 Probability: 0.971 AA1: | MKIKFKLSIVVIAIMAAVTA<br>MTKQKFLQPMWLELIIMLTAAAISIIGFSIISFLS |
| 84496 | 69 AA2: 70 Probability: 1.000 AA1: | RPHASTVMNYGMIALWLLAFFLIGFSIALVSVLA |
| 84511 | 27 AA2: 28 Probability: 0.963 AA1: | MGLINRYWRALSLAVAATALYAFNAAA |
| 84513 | 30 AA2: 31 Probability: 0.878 AA1: | MKLFHSLRFRFAVLLSVFIIALIGIVSVLG |
| 84514 | 37 AA2: 38 Probability: 1.000 AA1: | MSARAIIIFIVIFTAVAVMVVFLSFFLSLLISRPIQK<br>MNRKDIKNFTTKHTNQTNSYLRYGQENFTSNYA |
| 84517 | 74 AA2: 75 Probability: 0.610 AA1: | NNTNRNKLKRTTTRFELKFRVGSCVSWLIFFFFLFSFQAYA |
| 84527 | 20 AA2: 21 Probability: 0.876 AA1: | EATEVTIALSGANASSFSLS |
| 84545 | 31 AA2: 32 | MDDLRRKRSLDFILFLTALTLWVIGIFFVYS |
| 84554 | Probability: | MKLNAKLFVFGACFLMARPLFA |

| | | |
|---|---|---|
| | 1.000 AA1: 22 AA2: 23 Probability: 0.651 AA1: | |
| 84564 | 19 AA2: 20 Probability: 0.998 AA1: | ALWPRIGAVSTAAAQVLVA |
| 84565 | 28 AA2: 29 Probability: 1.000 AA1: | MIFGNKTTFIALVAIFGFALTFTFSCSG |
| 84579 | 20 AA2: 21 Probability: 0.676 AA1: | MKKIITALIVILVCAGAGFA |
| 84591 | 26 AA2: 27 Probability: 0.994 AA1: | MTTCIIKLWKRIIIAAVFLLFNYYSG |
| 84594 | 25 AA2: 26 Probability: 0.714 AA1: | MRKLILTTAAVAAAIGLAGCSKKPA |
| 84601 | 18 AA2: 19 Probability: 1.000 AA1: | GLSRIRFSGFTISTAVCA |
| 84609 | 23 AA2: 24 Probability: 0.895 AA1: | MKTKAIKFCLTLLIIFFVSPVFA |
| 84613 | 25 AA2: 26 Probability: 0.928 AA1: | MIKTVLGFGYLVFAMIFLLPFAITA |
| 84617 | 33 AA2: 34 Probability: 0.999 AA1: | MFPSQSFKSSYKYFDVLMTFFVVVLVLSNIASS |
| 84619 | 20 AA2: 21 Probability: 0.998 AA1: | MKNIAAMIVLAAMCSVGGCA |
| 84629 | 24 AA2: 25 Probability: 0.999 AA1: | MLSYMRPYAALIVTAFALALFAVA<br>MRNGIKKSEWRLIKTVIKEKFLNPLFIRRAALCLLG |
| 84659 | 47 AA2: 48 Probability: 0.987 AA1: | NFVLALGVAMA |
| 84691 | 28 AA2: 29 Probability: 0.852 AA1: | MKINSGFTTKYLSVIAATLLFFIAFVTG |
| 84704 | 22 AA2: 23 Probability: 0.971 AA1: | SLSRASFSRNLSFGAAPVLVLG |
| 84716 | 23 AA2: 24 Probability: 0.998 AA1: | MKIKFAPLIAPLFMLILFFSCLS |
| 84722 | 36 AA2: 37 | MVRTKEEVSSKNLLRSLLVGTLGSSVLILIAVAVMA |
| 84727 | Probability: | LEARQLIAVRRYSAALASLAIFQPVAV |

| | | |
|---|---|---|
| | 0.898 AA1: 27 AA2: 28 Probability: 0.697 AA1: | |
| 84774 | 30 AA2: 31 Probability: 0.757 AA1: | MIKLSAIFIAFFRVGAFTLGGGYAMFSVLR |
| 84784 | 22 AA2: 23 Probability: 0.974 AA1: | DYGEPRVAGIIAFVAIIGFSMA |
| 84911 | 27 AA2: 28 Probability: 1.000 AA1: | MSGGKISGTKASSSFSGGVAVGGVAHA |
| 84935 | 28 AA2: 29 Probability: 0.996 AA1: | MKKRLIFTFIVIASVLLFLAASPWEGSA |
| 84973 | 25 AA2: 26 Probability: 0.504 AA1: | MFSSVDTMWVLVAAALVFFMQAGFA |
| 84975 | 16 AA2: 17 Probability: 0.962 AA1: | MKRVVITGGSIVSALG |
| 84996 | 21 AA2: 22 Probability: 0.998 AA1: | IGESYTICAAYALFLAQSAFG |
| 84999 | 20 AA2: 21 Probability: 0.559 AA1: | MKKYIKYLLLPIIATLLVLA |
| 85005 | 23 AA2: 24 Probability: 0.990 AA1: | MNTKGTTGGSGFLSGVFNSPVLQ |
| 85017 | 21 AA2: 22 Probability: 0.999 AA1: | MKHLLLPLLCLLLFFCAKQQV |
| 85028 | 19 AA2: 20 Probability: 0.999 AA1: | MSFSRIALILLAAVSAIFA |
| 85058 | 46 AA2: 47 Probability: 0.718 AA1: | MKNTKKVFKGFSKLFWVIAITAMIVSVLT GCVSYSVQSVVISAANN |
| 85060 | 17 AA2: 18 Probability: 0.794 AA1: | MNGLFLLVISVVVLSLA |
| 85099 | 23 AA2: 24 Probability: 0.748 AA1: | MSQQPQKHALVTGATSGIGLAIA MQIKKILFVNKRLKTASDGVYALALLLPYGL |
| 85105 | 44 AA2: 45 Probability: 0.851 AA1: | LFSVFIAIPIAVA |
| 85109 | 22 AA2: 23 | VYYNNEGASTAAGSGSPTGILA |
| 85138 | Probability: | MFLRSLSVPVNTISGVGSSAA |

| | | |
|---|---|---|
| | 0.783 AA1: 21 AA2: 22 Probability: 1.000 AA1: | |
| 85171 | 20 AA2: 21 Probability: 0.986 AA1: | MRKLILTTAAVTAAIAGAFA |
| 85181 | 19 AA2: 20 Probability: 0.540 AA1: | GILRVLSGSSGGSVAAVVA |
| 85184 | 13 AA2: 14 Probability: 0.807 AA1: | MALTLSALGIVLA |
| 85186 | 28 AA2: 29 Probability: 1.000 AA1: | MNIGIIGLGLIGGSFAKAYKAAGHPVFA |
| 85190 | 23 AA2: 24 Probability: 1.000 AA1: | MSNIGKITAFVIFICLFPLPVFA |
| 85224 | 21 AA2: 22 Probability: 0.997 AA1: | MRKIITKAIALLIAFSTVTLA |
| 85233 | 38 AA2: 39 Probability: 0.993 AA1: | FNRSITGTFSRSIITTFTTMLAAISLFFLLPGSMKDFA |
| 85234 | 31 AA2: 32 Probability: 0.540 AA1: | MKRIIRFSKYFLPAAVLSMTIAVFSIVGFLT |
| 85245 | 35 AA2: 36 Probability: 0.949 AA1: | AYRELAVLRLYLPNGLFFWTSPFFSGGRFAGALMA |
| 85253 | 16 AA2: 17 Probability: 0.977 AA1: | MKPFFALIFMVTLVNA |
| 85273 | 23 AA2: 24 Probability: 0.961 AA1: | MKNIKDKLILITGGASGIGQLMA |
| 85280 | 35 AA2: 36 Probability: 1.000 AA1: | MGVYMRLFTKPRLTLLLATTIALFAICLSGCGGSG |
| 85312 | 29 AA2: 30 Probability: 1.000 AA1: | MTRFSFNRLGRALSVASAAVVGAVFSVGA |
| 85314 | 17 AA2: 18 Probability: 0.996 AA1: | VTKKTAACLLALPPVLA MINSKIKSLLSVPGSLSRLLNRRQKISFLILLILTIGYSIIE |
| 85376 | 56 AA2: 57 Probability: 0.989 AA1: | TVGISAIMPFISLA |
| 85406 | 20 AA2: 21 | MKTFRPRALLFFICLLLVCA |
| 85414 | Probability: | MGAAINIAVVICAAGASSRMK |

| | | |
|---|---|---|
| | 0.591 AA1: 21 AA2: 22 Probability: 1.000 AA1: | |
| 85472 | 33 AA2: 34 Probability: 0.535 AA1: | MYSKKNKNSILTYKMSTLKLTAALACTLLCVVA |
| 85473 | 25 AA2: 26 Probability: 0.795 AA1: | MNWLGALVIGLMQAVAIIPGVSRSG MLEKRKIKTNQVSGNAAMPRKKRRLKYIAVLPSFITLM |
| 85477 | 74 AA2: 75 Probability: 0.943 AA1: | NGVCGFFAIIIASRSPDIRWNFSFIPGAGMSPFTLA |
| 85482 | 27 AA2: 28 Probability: 0.989 AA1: | MFDFLRTLWNMLPSTLMMAAPIFIAAA |
| 85483 | 19 AA2: 20 Probability: 0.764 AA1: | MKINSVLVAGAGAVGLTVA |
| 85493 | 18 AA2: 19 Probability: 0.904 AA1: | ATHPALTSRLASTFSDAA |
| 85499 | 18 AA2: 19 Probability: 0.559 AA1: | GLSRALPGTATITATAQG |
| 85507 | 19 AA2: 20 Probability: 0.968 AA1: | VHAMEMGASAVLVNTAIAA |
| 85532 | 30 AA2: 31 Probability: 1.000 AA1: | MSSKQKIIMISAGAALAVLIVLGIIFFVKG |
| 85554 | 19 AA2: 20 Probability: 0.976 AA1: | MKRVVFMFAFILAGISAFA |
| 85558 | 29 AA2: 30 Probability: 0.998 AA1: | AYREFRHQKSRITQALMITGTSAAGAAAG |
| 85575 | 30 AA2: 31 Probability: 1.000 AA1: | MRKLILTTAAVAAAIGLAGCSKTPTPKSVS |
| 85619 | 22 AA2: 23 Probability: 0.910 AA1: | MKKIFKVLAIIAIVAVIGFSMA MLETYDKLLSSVLIGNNISNIASSTLATALFIGLFGVNGVSI |
| 85633 | 63 AA2: 64 Probability: 0.996 AA1: | ATLIMTILLMLFCDISPKALA |
| 85639 | 26 AA2: 27 Probability: 0.988 AA1: | MITKLIKIALMVSIALSVSFTLSCGQ MTVIAFTTTVIAFTTTVIAFTTTVIAFTTTVIAFTTTVIAF |
| 85640 | 68 AA2: 69 | TTTGIAFTTTVIAFTTTVIAFTTTVIA |
| 85646 | Probability: | MNNNDPILWKFLILVVLIFASAFFTLA |

| | | |
|---|---|---|
| | 0.998 AA1: 27 AA2: 28 Probability: | |
| | 0.992 AA1: | |
| 85654 | 24 AA2: 25 | MTCKRIAPAILAALLPLSLAVLMG |
| | Probability: 0.996 AA1: | |
| 85677 | 24 AA2: 25 | MDKLLLMILAVSFSACSSTHVRSG |
| | Probability: 0.749 AA1: | |
| 85696 | 26 AA2: 27 | MSIRIRAFLIIITTNLLIILFSVFLG |
| | Probability: 0.937 AA1: | |
| 85714 | 26 AA2: 27 | TSPVLIKGIFPTVLPALVPGATMAFA |
| | Probability: 0.891 AA1: | |
| 85720 | 26 AA2: 27 | MNGIKSIITNTMFVTGAALVILTVFS |
| | Probability: 0.698 AA1: | |
| 85726 | 29 AA2: 30 | MEKNAALAFRSAAVFIILFQIRLIAADLA |
| | Probability: 0.948 AA1: | |
| 85753 | 15 AA2: 16 | LPQSAGTAAATPVQA |
| | Probability: 0.999 AA1: | MIARRATLVAIALAWFLLGAVGFFSDAVSLV |
| 85756 | 44 AA2: 45 | WLFSGLLLLPVIA |
| | Probability: 0.967 AA1: | |
| 85769 | 33 AA2: 34 | MGTQKISVSRRRIRLIAWISGLAAFCLMFASLV |
| | Probability: 0.859 AA1: | |
| 85788 | 24 AA2: 25 | ITADSRGKRVLNFFCYTAAFSIHA |
| | Probability: 1.000 AA1: | |
| 85818 | 22 AA2: 23 | MKRSTLLCILSGVFLAATLAPA |
| | Probability: 1.000 AA1: | |
| 85823 | 21 AA2: 22 | MKKQIAVLLCVAALIIQPLFA |
| | Probability: 0.646 AA1: | |
| 85842 | 19 AA2: 20 | GLRAGPALAIMGFIVGAKA |
| | Probability: 0.709 AA1: | |
| 85858 | 18 AA2: 19 | MIIGVAGLGLIGGSLAKA |
| | Probability: 0.984 AA1: | VVVALFAVCGLIVTIGRKSRLNVFRSKMFF |
| 85891 | 42 AA2: 43 | MAFCVAALVAFA |
| | Probability: 0.997 AA1: | |
| 85902 | 23 AA2: 24 | MKIKIKLSIMMIAIVAIVAGGVA |
| 85914 | Probability: | VVSANPASSFITATSITIA |

| | | |
|---|---|---|
| | 0.591 AA1: 19 AA2: 20 Probability: 0.999 AA1: | |
| 85932 | 25 AA2: 26 Probability: 0.952 AA1: | LIGENERVVLCLLVLLVLAGGTVFA |
| 85966 | 22 AA2: 23 Probability: 1.000 AA1: | MKKIILLMAAVILNISGCKENA |
| 85969 | 29 AA2: 30 Probability: 0.978 AA1: | LKFFGHRGFKLCLFALLLAGLILAAPLYG |
| 85979 | 31 AA2: 32 Probability: 0.946 AA1: | MRCLFWANAIRPYTLAAGFVLAMVFTFSCSS |
| 85980 | 23 AA2: 24 Probability: 0.841 AA1: | RSEYEASVFVLVMVLITAAFVCA VESQTEDIVNTSRSAKKRGINLLSFSLLSG |
| 85996 | 39 AA2: 40 Probability: 0.840 AA1: | LLLSSVGIA MPEQKIAYRNKPQINLSYPSTGRMWLVS |
| 86005 | 41 AA2: 42 Probability: 0.997 AA1: | VCAFLCALQSALT |
| 86023 | 27 AA2: 28 Probability: 1.000 AA1: | MGFLKKQGRNICFAVILIALALMPVFS |
| 86043 | 29 AA2: 30 Probability: 0.570 AA1: | MINSRKIINIILSVLFISMTVSSCGKSTA |
| 86069 | 20 AA2: 21 Probability: 0.941 AA1: | TTYSGTFSVSGSTITFNATG |
| 86070 | 27 AA2: 28 Probability: 0.635 AA1: | MKFLHALEMAGIFLLFAAAAGPHFVFT |
| 86071 | 28 AA2: 29 Probability: 0.567 AA1: | CFRNTCLLAAKTGLTSVAFPNISTGVYG |
| 86097 | 24 AA2: 25 Probability: 0.802 AA1: | MLPRRLIDTVFTWLFRAFSFTGVA |
| 86114 | 31 AA2: 32 Probability: 0.636 AA1: | MAEGSASVLIIAAPLLISALVIGLVVAILQA |
| 86115 | 26 AA2: 27 Probability: 0.677 AA1: | MVDNAVLQNILLFGPVFLLIAARALA |
| 86119 | 23 AA2: 24 | MLPIICFISITLLFLDFTGTVHT |
| 86145 | Probability: | MRLEGKTALISGAAQGLGAAMARRFA |

| | | |
|---|---|---|
| | 0.793 AA1: 26 AA2: 27 Probability: 0.978 AA1: | |
| 86172 | 30 AA2: 31 Probability: 0.999 AA1: | MFTFLHKARKFFPAFILLFTSFFFPNGVKG |
| 86173 | 34 AA2: 35 Probability: 0.999 AA1: | DFRLAKKVATGTLKKAATTVSTPTAATPSPNAAA |
| 86179 | 21 AA2: 22 Probability: 0.655 AA1: | MKSITKLTATLALAITLTFTA |
| 86185 | 26 AA2: 27 Probability: 0.999 AA1: | ISGTADPASAWSIAVDSTLTAGSVLA |
| 86248 | 22 AA2: 23 Probability: 0.999 AA1: | MRAFRGFKIWVAVLLIATASWA |
| 86274 | 20 AA2: 21 Probability: 1.000 AA1: | MSLAHIKIGFALLLAGTAYA |
| 86356 | 23 AA2: 24 Probability: 0.626 AA1: | MKNTLWVIAILTAFTLSTTSCFA<br>MGNGKKSKDKFRLVDLLVILICLFGAFYSVN |
| 86367 | 43 AA2: 44 Probability: 1.000 AA1: | MFWSDLFQTLNG<br>MKLSQMSIKKKMLAFSGLFFLVVTVLGTAA |
| 86394 | 44 AA2: 45 Probability: 1.000 AA1: | FFFSMSQIVRSSAA |
| 86395 | 19 AA2: 20 Probability: 0.982 AA1: | MKNKVSLFLLCAIASGLWA |
| 86397 | 26 AA2: 27 Probability: 0.992 AA1: | MEKIKLSLGILAVTGLVSCASGPRIS |
| 86411 | 36 AA2: 37 Probability: 1.000 AA1: | MKKTLRFSKFFLPAAIISAIIVILSISGYFINKGFA<br>MLNQKKDPILIVKVCLLILIGAFAFSFISSLFN |
| 86412 | 38 AA2: 39 Probability: 0.999 AA1: | ARAMA |
| 86418 | 26 AA2: 27 Probability: 0.999 AA1: | MSMRRFSICLCLIAVLFLSCNGNKVA |
| 86432 | 23 AA2: 24 Probability: 0.762 AA1: | MTIFCKKQRILAVLLVFLFSACA |
| 86444 | 20 AA2: 21 | MKRLFVFPAWILIFSLTIFS |
| 86494 | Probability: | MKKTIFAFSFLFAAFLYA |

| | | |
|---|---|---|
| | 1.000 AA1: 18 AA2: 19 Probability: 0.664 AA1: | |
| 86528 | 30 AA2: 31 Probability: 0.994 AA1: | MLNNLFSIDNALLILRLATPLILVSLSSTV |
| 86540 | 30 AA2: 31 Probability: 0.543 AA1: | MFVRFHFVPLKAVLLFFLTSLSGYATPVVS |
| 86542 | 17 AA2: 18 Probability: 0.985 AA1: | MFTGGTALFTLSVRALA |
| 86551 | 28 AA2: 29 Probability: 0.999 AA1: | MEEKMSRKSSFKSVVVVILTAFILAAAG |
| 86560 | 23 AA2: 24 Probability: 0.590 AA1: | MKNKQKMFGIIAASVAIIGFSFA<br>AWLFGSYSEKTFANISGKPNKSVAGFIGG |
| 86561 | 39 AA2: 40 Probability: 1.000 AA1: | ALGSVMVAVG |
| 86573 | 30 AA2: 31 Probability: 0.766 AA1: | MIYSLRTKILAIGLAFFALIGTAFVLYSMA<br>LLALRFFFPFRRNIPNSKAVIVRTLLKEAPV |
| 86601 | 69 AA2: 70 Probability: 1.000 AA1: | LGACGLGYAATMILFVIANKLTASANAIMLQYTAPVWA |
| 86602 | 27 AA2: 28 Probability: 0.937 AA1: | MQIKKPFKHIALIAAITAILAPRPALS |
| 86636 | 42 AA2: 43 Probability: 0.521 AA1: | MQTRKSTKSAPQYKTKSNFRFADFFIIVFFLSLAFISIILFR |
| 86644 | 20 AA2: 21 Probability: 0.976 AA1: | MKKILLTGGGTAGHVTPNLA |
| 86654 | 23 AA2: 24 Probability: 0.962 AA1: | MAKKKIALVGAGQIGGTMALVLA<br>MRKTYNYSAKLALLHRFWLIVASILLSSVLLL |
| 86663 | 42 AA2: 43 Probability: 0.999 AA1: | MFILFQKSHR |
| 86693 | 27 AA2: 28 Probability: 0.512 AA1: | MKMKLTAAIIAAFVIIMGCAGGPKASK |
| 86714 | 29 AA2: 30 Probability: 0.863 AA1: | MTWKKTKEWEYFPFVLLAFMVLIMHILMK |
| 86735 | 35 AA2: 36 | MKETQDNEKKTSIKVRYIIFSISLFLIIFIGGSAA |
| 86749 | Probability: | LIGNTIVNVVTSALATMLFMGLFGA |

| | | |
|---|---|---|
| | 0.891 AA1: | |
| | 25 AA2: 26 | |
| | Probability: | |
| | 0.755 AA1: | |
| 86750 | 30 AA2: 31 | MAQDILSKQVPLGQVVLLVLFSLPSVIALS |
| | Probability: | |
| | 0.831 AA1: | MVVQYAIFWHITLKSKSGGMMTLFALIGLLP |
| 86771 | 43 AA2: 44 | MCVISPFAGVWA |
| | Probability: | |
| | 1.000 AA1: | |
| 86775 | 29 AA2: 30 | MKIFNRQGRALLLTAAAMVATGFASKVSA |
| | Probability: | |
| | 0.991 AA1: | |
| 86783 | 37 AA2: 38 | MNIKQSVNNFKTRVLSSYIMIPAVILALLLIAAISNN |
| | Probability: | |
| | 0.977 AA1: | |
| 86786 | 27 AA2: 28 | MKMKKICIFCSLSSFGTGMPISTYKLA |
| | Probability: | |
| | 0.969 AA1: | |
| 86807 | 30 AA2: 31 | MSRKQKIIIISAGVALTALIVLGIIFFVKG |
| | Probability: | |
| | 0.835 AA1: | |
| 86837 | 22 AA2: 23 | VASKETTTIVALATPPGTSAIA |
| | Probability: | |
| | 0.618 AA1: | |
| 86847 | 35 AA2: 36 | VKLRSILEIDYLLLLPVIVLMMLGVFFIYSSGITS |
| | Probability: | |
| | 0.767 AA1: | |
| 86906 | 39 AA2: 40 | LSFKETFSLIKIYALSKIKPQFKAVLMVSIFLAAVQYFV |
| | Probability: | |
| | 0.586 AA1: | |
| 86910 | 21 AA2: 22 | MIMNALLSAKATASLSRFRAA |
| | Probability: | |
| | 0.927 AA1: | |
| 86914 | 33 AA2: 34 | MVISLLATWKKRGFTQTLSASKLLLTPSTSSLS |
| | Probability: | |
| | 0.772 AA1: | |
| 86915 | 17 AA2: 18 | MLFIFFVSFLFAFSYGA |
| | Probability: | |
| | 0.997 AA1: | |
| 86927 | 25 AA2: 26 | MKTTRNRLVFAAVFAALIAVSNFFQ |
| | Probability: | |
| | 0.784 AA1: | |
| 86932 | 22 AA2: 23 | MAGVCIGTFALIVALSAVNGFE |
| | Probability: | |
| | 0.841 AA1: | |
| 86937 | 17 AA2: 18 | MKRFFILLFFIPVVSCA |
| | Probability: | |
| | 1.000 AA1: | MKNVFKIKMFGNLNRDRRSKVPLLIIALLAII |
| 86939 | 40 AA2: 41 | MFSLATCS |
| 86951 | Probability: | MSKVLHILSIFIILLAIIASGFGLFYK |

| | | |
|---|---|---|
| | 0.790 AA1: | |
| | 27 AA2: 28 | |
| | Probability: | |
| | 1.000 AA1: | MNIKQAVKTIKVRVLNSYILVLVVVLAFALMV |
| 86956 | 53 AA2: 54 | ALSSYFMTDIVKRQLASNAHA |
| | Probability: | |
| | 1.000 AA1: | |
| 86965 | 24 AA2: 25 | MNRSKSKLLTVVAAIFAIAAYSFA |
| | Probability: | |
| | 0.743 AA1: | |
| 86976 | 26 AA2: 27 | GQSIAREWRKAFGFAGILNISAKQMA |
| | Probability: | |
| | 0.999 AA1: | |
| 86981 | 27 AA2: 28 | LSTFIKPLVALLISILLFAGLIFLANG |
| | Probability: | |
| | 0.985 AA1: | |
| 86982 | 19 AA2: 20 | MKRIIFLLTLTAAVLFTLG |
| | Probability: | |
| | 0.839 AA1: | |
| 87027 | 36 AA2: 37 | AGQYAKPDFNLLTFSLLSGLLLSSVGIALTIFLSIA |
| | Probability: | |
| | 0.570 AA1: | TLTGSPISSIAAGGTGTFSVRPNASLIAGTYTATV |
| 87069 | 43 AA2: 44 | TVSGGSVA |
| | Probability: | |
| | 0.997 AA1: | KRLIRKNGKNRRHNRRFSVASTLTTALVFVPVIFL |
| 87115 | 44 AA2: 45 | PGPLGSLFG |
| | Probability: | |
| | 0.935 AA1: | |
| 87128 | 32 AA2: 33 | VRPPFPFRSVAKSSSSSVLSSSSSKQSSSSFT |
| | Probability: | |
| | 0.651 AA1: | |
| 87179 | 33 AA2: 34 | YLTIPGFFNAYKNTIFYTVVGTFISLSVLLLFA |
| | Probability: | |
| | 0.647 AA1: | |
| 87208 | 20 AA2: 21 | RRGDLTALTGSLPWISLALA |
| | Probability: | |
| | 0.993 AA1: | VHRPTINGKGFIALAAVSFGRWVPLGILGSSFL |
| 87209 | 41 AA2: 42 | FGMSSALA |
| | Probability: | |
| | 0.920 AA1: | |
| 87266 | 20 AA2: 21 | VNKKYDYLIVGAGLFGAVFA |
| | Probability: | |
| | 0.831 AA1: | |
| 87281 | 19 AA2: 20 | RFCLVMLVIVLAFGMTVVG |
| | Probability: | |
| | 0.995 AA1: | |
| 87290 | 28 AA2: 29 | MKVKYFYLTIITGIAVTLLSSCSSAKFG |
| | Probability: | |
| | 0.820 AA1: | |
| 87346 | 32 AA2: 33 | YQWYSNTNNNKIGGTLLSSATNTSYTPLNTAA |
| 87423 | Probability: | MKKGSIFIFILYTFLYVSGSVHG |

| | | |
|---|---|---|
| | 0.961 AA1: 23 AA2: 24 Probability: | |
| | 0.643 AA1: | |
| 87446 | 28 AA2: 29 | MQKMYKYPWLIIAVVAVITVFFGFQLPK |
| | Probability: 0.763 AA1: | |
| 87448 | 26 AA2: 27 | MGDIKKTNPILLVLIYILMVVFAVLA |
| | Probability: 0.959 AA1: | MKETQDNEKKTSIKVRYIIFSISLFLIIFIGG |
| 87467 | 49 AA2: 50 | SAAFFFSMWKMLHSTAG |
| | Probability: 1.000 AA1: | |
| 87473 | 22 AA2: 23 | MKLNTRKLIAIFLAVFAITTYA |
| | Probability: 0.826 AA1: | |
| 87502 | 23 AA2: 24 | VKQYDLVVVGGGPAGLAAAIAAA |
| | Probability: 1.000 AA1: | |
| 87519 | 20 AA2: 21 | MISRLFFAVFLTILSAITSA |
| | Probability: 0.995 AA1: | |
| 87527 | 29 AA2: 30 | MKKPKKTSFAKLFTTVSLLMILTITVSLS |
| | Probability: 0.981 AA1: | |
| 87528 | 22 AA2: 23 | MNKKTVLCILILIMVYASCSQK |
| | Probability: 0.999 AA1: | |
| 87543 | 22 AA2: 23 | MIRLNKLFFVYFFFLAVFTVTA |
| | Probability: 1.000 AA1: | |
| 87556 | 22 AA2: 23 | MKRYLFISLLLITVLACTGAQQ |
| | Probability: 1.000 AA1: | |
| 87557 | 21 AA2: 22 | MKRKVFLAVTLFAFMAVGVYA |
| | Probability: 1.000 AA1: | |
| 87567 | 19 AA2: 20 | MKAFIAAALLILIIGNIQA |
| | Probability: 0.726 AA1: | |
| 87573 | 30 AA2: 31 | MFSILHKTVKFFPVLVLLIFSLCFPNYIKG |
| | Probability: 0.950 AA1: | |
| 87582 | 20 AA2: 21 | MKCFTKILILFLLLSLVSCK |
| | Probability: 0.803 AA1: | |
| 87583 | 20 AA2: 21 | MRCFLKIFTIFLLFSLFSCK |
| | Probability: 0.831 AA1: | |
| 87602 | 17 AA2: 18 | MNGLLLLLISVITLSLA |
| 87620 | Probability: | MKKVVLLAALIICVMLTVSCKS |

| | | |
|---|---|---|
| | 0.994 AA1: 22 AA2: 23 Probability: 0.559 AA1: | |
| 87623 | 17 AA2: 18 Probability: 0.909 AA1: | MNGLFLLVISVVVLSLA |
| 87638 | 19 AA2: 20 Probability: 0.943 AA1: | MTRKNLPVILIFTVFSIWA |
| 87651 | 18 AA2: 19 Probability: 0.831 AA1: | MTRKQASTLSVMILAVIA |
| 87672 | 22 AA2: 23 Probability: 0.849 AA1: | MPWRLILFIVIFAIFLAFITVN |
| 87674 | 23 AA2: 24 Probability: 0.782 AA1: | MKKRKLASAIGYKSGDFAPALLA |
| 87676 | 20 AA2: 21 Probability: 0.988 AA1: | MSGKRISILIGIAIPMILFS |
| 87680 | 21 AA2: 22 Probability: 0.998 AA1: | MKRRIAIIIIITVLTVTAVA |
| 87699 | 19 AA2: 20 Probability: 0.971 AA1: | MKKRVLIFLLLNVSCIIYA |
| 87710 | 29 AA2: 30 Probability: 0.525 AA1: | MKKLPEKALFIIFFALAFSQFPKNALLHA |
| 87717 | 17 AA2: 18 Probability: 0.999 AA1: | VIVFLTVLTFTASRAES |
| 87738 | 26 AA2: 27 Probability: 0.967 AA1: | MSMRRFSICLCLIAVLFLSCNGNKVA |
| 87746 | 33 AA2: 34 Probability: 0.778 AA1: | MGTQKISVSRRRIRLIAWISGLAAFCLMFASLV |
| 87769 | 21 AA2: 22 Probability: 0.998 AA1: | MSSPDIRAGMAMLIAAMTAQG |
| 87799 | 19 AA2: 20 Probability: 1.000 AA1: | MKTAIFILLCIICINGVWA |
| 87802 | 18 AA2: 19 Probability: 1.000 AA1: | MKKLLVIALFLLPLAIWA |
| 87807 | 34 AA2: 35 | MPIPTSKRGHPLFSTLAAVAFALLAAASAAPLFA |
| 87810 | Probability: | MPKRNDIKKILLIGSGPIVIGQA |

| | | |
|---|---|---|
| | 0.578 AA1: 23 AA2: 24 Probability: 0.965 AA1: | |
| 87813 | 23 AA2: 24 Probability: 0.970 AA1: | MTKHFFSFIFFFFLFYSILPAQE |
| 87816 | 37 AA2: 38 Probability: 1.000 AA1: | MRYRLRSKLALFTIALVLLVDLLISAPLFFMIIKTQQ |
| 87845 | 26 AA2: 27 Probability: 1.000 AA1: | MKKSNVKSLAVLLLFIIFGTPLSALG |
| 87848 | 21 AA2: 22 Probability: 1.000 AA1: | MKKQIAVLLCVAALIIQPLFA |
| 87855 | 23 AA2: 24 Probability: 0.946 AA1: | LTLALSAFFAVTAAAPAAVPVAA |
| 87860 | 19 AA2: 20 Probability: 0.990 AA1: | MIIIVITAVFALAIAFVLG |
| 87870 | 21 AA2: 22 Probability: 0.653 AA1: | MKHLLLPLLCLLLFFCAKQQV |
| 87921 | 22 AA2: 23 Probability: 1.000 AA1: | MRFFHILIICLLLLPACKKREA |
| 87926 | 22 AA2: 23 Probability: 0.991 AA1: | MNKKTEFLILLLIFIVSSSLFA |
| 87929 | 25 AA2: 26 Probability: 0.994 AA1: | MIKTILGFGWLIFAMIFLLPFALTA |
| 87932 | 21 AA2: 22 Probability: 0.944 AA1: | MKPLMCLIVVMFLLPCGGINA<br>MTERELILKQNPLVNVSRSTAARMWLVAIC |
| 87954 | 41 AA2: 42 Probability: 0.911 AA1: | AFLAALQSAMG |
| 87973 | 23 AA2: 24 Probability: 0.999 AA1: | MLRPEAAVSGRVLAAADSSAVSA |
| 87985 | 24 AA2: 25 Probability: 1.000 AA1: | MKIQTRLRFLALGIVLVSLISCQA |
| 87987 | 22 AA2: 23 Probability: 1.000 AA1: | MRRRIIFSALLLLALVPQAAYT |
| 87989 | 19 AA2: 20 | MKTISIILGLLICSGPALA |
| 87994 | Probability: | MKNKLWVIVILTAFTLSATSCSA |

| | | |
|---|---|---|
| | 1.000 AA1: 23 AA2: 24 Probability: 0.993 AA1: | |
| 87997 | 26 AA2: 27 | MNKIASIKLPIIALAMVFTISCSSDS |
| | Probability: 0.698 AA1: | |
| 88022 | 32 AA2: 33 | MTKTKGEIPVKLWDMVIILLALCMTGFSFFSA |
| | Probability: 0.999 AA1: | |
| 88029 | 20 AA2: 21 | MKNIAVLSVLVIILSVNSYA |
| | Probability: 0.623 AA1: | NAPVEWGIYRNIPFAILGFIIIIIWFRMTKKTN |
| 88032 | 58 AA2: 59 | DRVFRFLWLAVLLSFAFYIPVVLWA |
| | Probability: 0.991 AA1: | |
| 88038 | 28 AA2: 29 | MRKLILTTAAVAAAIGLAGCGRTPTSES |
| | Probability: 1.000 AA1: | |
| 88039 | 27 AA2: 28 | MQIKKPFKHIALIAAITAILAPRPALS |
| | Probability: 0.800 AA1: | |
| 88070 | 36 AA2: 37 | FWRLIGVCVVLLSNIIVIKISAFYLSLLFFTGQVFS |
| | Probability: 0.526 AA1: | |
| 88071 | 23 AA2: 24 | MTIKPNDGRSFIPTYALTFTAFA |
| | Probability: 0.997 AA1: | |
| 88086 | 20 AA2: 21 | SRIHIGAGASLISCPRRATA |
| | Probability: 0.996 AA1: | |
| 88095 | 32 AA2: 33 | MKIKIKLSLIMIAIMAIVAGGITVIELMQASA |
| | Probability: 0.961 AA1: | |
| 88110 | 20 AA2: 21 | MKLSIRVPLLITAVVLITSA |
| | Probability: 1.000 AA1: | |
| 88115 | 18 AA2: 19 | VVLCLLVLLVLAGGTVFA |
| | Probability: 0.994 AA1: | |
| 88134 | 21 AA2: 22 | MRQKLCLLACIVLLCGCSRNG |
| | Probability: 0.907 AA1: | |
| 88151 | 32 AA2: 33 | MKNGKMIVRVFAAVLSLIAIIGMIGCGGKSDT |
| | Probability: 0.959 AA1: | |
| 88153 | 30 AA2: 31 | GQRYKAYALSVLFLIGSFIAMIVGGGPRMG |
| | Probability: 0.811 AA1: | LLKEAPVLGACGFGYAATMILFVIANKLTASAN |
| 88159 | 45 AA2: 46 | AIMLQYTAPVWA |
| 88163 | Probability: | LIQNHWKKLKSFCALFIIIQFTSLYA |

| | | |
|---|---|---|
| | 0.998 AA1: 26 AA2: 27 Probability: | |
| | 0.999 AA1: | |
| 88168 | 23 AA2: 24 | MRKHILCLMATIFALAFTPIACG |
| | Probability: 0.979 AA1: | |
| 88169 | 19 AA2: 20 | ALTAASVSAGFFHSVATMA |
| | Probability: 0.977 AA1: | |
| 88172 | 24 AA2: 25 | MELGVMILIAALILLAIIIFASVA |
| | Probability: 0.999 AA1: | |
| 88180 | 21 AA2: 22 | MQKAAALFLILFFGVFVLLFA |
| | Probability: 0.998 AA1: | |
| 88199 | 28 AA2: 29 | MLRNFSIGVRLLFIIFILLLTIAALVFA |
| | Probability: 0.981 AA1: | MAIQFLIIIAVILILLLCIKMKTKRQRIAVLI |
| 88216 | 46 AA2: 47 | GFVVSILSFFAFLA |
| | Probability: 0.700 AA1: | |
| 88222 | 24 AA2: 25 | MSPIFVVFIAVAAVAAFIGGFFWR |
| | Probability: 1.000 AA1: | |
| 88224 | 31 AA2: 32 | MKVFNRQGRGRTLLLAAVVAVGAALVSGAVA |
| | Probability: 0.902 AA1: | |
| 88233 | 40 AA2: 41 | VAMTIRSFFEKNRKGISYYLKLLVVVVAFVIIIVSSCLYA |
| | Probability: 0.995 AA1: | |
| 88236 | 27 AA2: 28 | MKKQFSLVFFTLLIVCFLGALPFYFSG |
| | Probability: 1.000 AA1: | |
| 88260 | 31 AA2: 32 | MIINPVKSIKKFVNKTFFIFSCILFFIIALA |
| | Probability: 0.995 AA1: | |
| 88263 | 33 AA2: 34 | MPSLQKKQKRIRFILLIVVATVLIINVASIVVG |
| | Probability: 0.984 AA1: | |
| 88273 | 42 AA2: 43 | VVVALFAVCGLIVTIGRKSRLNVFRSKMFFMAFCVAALVAFA |
| | Probability: 0.591 AA1: | |
| 88276 | 32 AA2: 33 | MTILTLLLRMTGGLCMFLFGMKMMSDGLQKSA |
| | Probability: 0.892 AA1: | |
| 88281 | 29 AA2: 30 | MSAIKRSFMLVLYSCLSLVSMPYWPMDYA |
| | Probability: 0.739 AA1: | |
| 88323 | 35 AA2: 36 | MNDNENLKPALSKWLTKIALIIFFAALTTTGAYIA |
| 88328 | Probability: | MQQLYFALAIMIGLIVLSILVFV |

| | | |
|---|---|---|
| | 0.791 AA1: 23 AA2: 24 Probability: | |
| | 0.732 AA1: | |
| 88347 | 19 AA2: 20 | MAGLLLIGGFAFRFLPVSA |
| | Probability: 0.617 AA1: | |
| 88361 | 21 AA2: 22 | MLLSVFVITIAGSFLALFLLS |
| | Probability: 0.711 AA1: | |
| 88363 | 31 AA2: 32 | MKKFLNEFKQFALKGNVLNLAVGVIIGAAFQ |
| | Probability: 0.997 AA1: | |
| 88368 | 25 AA2: 26 | MKGIYRFSAIIIAALIAIPAGYGQS |
| | Probability: 0.949 AA1: | |
| 88383 | 31 AA2: 32 | MKNILRKNYVFVATAILIVSAYFISGVLIYA |
| | Probability: 0.967 AA1: | |
| 88388 | 29 AA2: 30 | MAKYILIRLAQAALVLLLATIVTFGLVIV |
| | Probability: 0.988 AA1: | |
| 88394 | 28 AA2: 29 | MAKTSRKNSFIRVALVAFAVSSLLFLTS |
| | Probability: 0.997 AA1: | |
| 88404 | 19 AA2: 20 | MQKLLAVFVVLLFALVGLG |
| | Probability: 0.753 AA1: | |
| 88425 | 23 AA2: 24 | MNRKSAKKNLFITAPGGYGKTTA |
| | Probability: 0.997 AA1: | |
| 88426 | 25 AA2: 26 | MKTLFGKLLISFGVIILLMIVSVLA |
| | Probability: 0.999 AA1: | |
| 88465 | 18 AA2: 19 | MNKIATLLLLCTLLCVVA |
| | Probability: 1.000 AA1: | |
| 88469 | 31 AA2: 32 | MKSKKIAALFLTAAVLCAVLVTTGCSVRAEA |
| | Probability: 0.985 AA1: | |
| 88478 | 29 AA2: 30 | MILNIFGVNTASLIALLGAAGLAIGLALK |
| | Probability: 0.932 AA1: | |
| 88496 | 26 AA2: 27 | MGFFDRQGRGRALLTVAAAVCFGLTA |
| | Probability: 0.796 AA1: | |
| 88506 | 24 AA2: 25 | MTDSRVFRRYAGGLFFLAAAVFCA |
| | Probability: 1.000 AA1: | |
| 88541 | 21 AA2: 22 | MSKKIFLSFWFVFCLSFTAAA |
| 88568 | Probability: | MKKFCAALFLLTAALSFA |

| | | |
|---|---|---|
| | 1.000 AA1: 18 AA2: 19 Probability: 1.000 AA1: | |
| 88575 | 22 AA2: 23 Probability: 0.985 AA1: | MLLRKIILIISLIFFISAALWA |
| 88584 | 28 AA2: 29 Probability: 0.999 AA1: | MEEKMSRKSSFKSVVVVILTAFILAAAG |
| 88589 | 19 AA2: 20 Probability: 0.988 AA1: | MKKFLASYSLLILFGLAGA |
| 88632 | 23 AA2: 24 Probability: 0.962 AA1: | MPKFLLSLLLTALLISCGQKEQS |
| 88651 | 18 AA2: 19 Probability: 1.000 AA1: | MKLIPLVLIAVLPCACFG |
| 88656 | 32 AA2: 33 Probability: 0.982 AA1: | MRGLGRTVFKRRYRIAALAAALAIGAVNSANA |
| 88668 | 42 AA2: 43 Probability: 0.930 AA1: | MTGIRYILRRIFYGKDGRISIPKTVSFVLAAPPLLFVLAAFA |
| 88670 | 27 AA2: 28 Probability: 0.970 AA1: | VPESPVKNLDRGFLAVGLVAFIFSLLA |
| 88674 | 22 AA2: 23 Probability: 0.995 AA1: | MKNAIKIAVLTVVLAAGCMSYK |
| 88718 | 28 AA2: 29 Probability: 1.000 AA1: | MKRVITVLVALLMLAVVFVGCKPTSVTT |
| 88720 | 20 AA2: 21 Probability: 0.902 AA1: | MIIRAATIALLASISPVAQA |
| 88725 | 27 AA2: 28 Probability: 1.000 AA1: | VGIDMKKNNKNAFGSPLPAALSTSIAA |
| 88732 | 21 AA2: 22 Probability: 0.890 AA1: | MKKKNGILVLLLTGLCVTAFA |
| 88736 | 24 AA2: 25 Probability: 0.924 AA1: | MKILKYVIGLLLAAIGLYIFFRGS |
| 88739 | 35 AA2: 36 Probability: 0.512 AA1: | MKTKGNFLIKFIQAVGFAVIAAVFLYVYFALPDTA |
| 88741 | 30 AA2: 31 | MDLNVLLGIGAGIAVFTGIGAGIGIGIATA |
| 88743 | Probability: | MRKGLSFCIIFLAAFVFIYA |

| | | |
|---|---|---|
| | 0.999 AA1: 20 AA2: 21 Probability: 1.000 AA1: | |
| 88746 | 30 AA2: 31 Probability: 0.915 AA1: | MRSLRTKILTIVLAFLMFVSAAFVIYSLFA |
| 88761 | 27 AA2: 28 Probability: 0.820 AA1: | MRYRYNAIIGFTAALLFSLLYIIGAFN |
| 88770 | 31 AA2: 32 Probability: 0.999 AA1: | VLKLAKYLKPFIPLIILAVAVLFLQAFCDLS MKKPKTSLRKMVWGEHKISLRFRFINIALFL |
| 88774 | 43 AA2: 44 Probability: 1.000 AA1: | LTSLIMAVVLLS |
| 88775 | 38 AA2: 39 Probability: 0.919 AA1: | MMKEKQKQERRTSIKKRFIIFSAVLFLVIFIGGNAAFA |
| 88778 | 23 AA2: 24 Probability: 1.000 AA1: | MVWRMNKKALIAMSGGVDSSVAA |
| 88793 | 20 AA2: 21 Probability: 0.825 AA1: | MRKLILTTAAVAAAIAGAFA |
| 88820 | 21 AA2: 22 Probability: 0.823 AA1: | VLNIAGTVYICLFAGPGSVMN MDVNMLIYFGLAVLAVIAVVLVYFLASFFSLW |
| 88850 | 40 AA2: 41 Probability: 0.970 AA1: | LQALFSKA MKNKLSGVIHPAVIAVWAAVCAAGFLLPTVPI |
| 88853 | 46 AA2: 47 Probability: 0.848 AA1: | WGTGSVFSVSSALA MSNSRKFKERRDGLKVRFSIGAKLITIISIIVLISL |
| 88857 | 43 AA2: 44 Probability: 0.999 AA1: | GSITALV |
| 88870 | 19 AA2: 20 Probability: 0.994 AA1: | MNKTHFALMLCLATSIVCA |
| 88890 | 21 AA2: 22 Probability: 0.622 AA1: | MRIKTFITLAVFTGLAVSITA |
| 88902 | 21 AA2: 22 Probability: 0.971 AA1: | GPFLRFPGPLTIAVLPGLPHS |
| 88912 | 23 AA2: 24 Probability: 0.743 AA1: | MKPMNKIAISLTLLALLIGCYKA |
| 88924 | 22 AA2: 23 | MFAGSLIMMIYNSWMLALVAVA |
| 88927 | Probability: | MNNAMSSYSFHGIFCVPVRKILFI |

| | | |
|---|---|---|
| | 0.751 AA1: 40 AA2: 41 Probability: | LFFYSSFLIPNSSLFA |
| | 0.732 AA1: | |
| 88940 | 21 AA2: 22 Probability: | MKELQYDAVVIGAGAAGMAAA |
| | 0.999 AA1: | |
| 88963 | 22 AA2: 23 Probability: | VKNLRPLVIFIFCTLAAAQLQA |
| | 0.981 AA1: | |
| 88985 | 29 AA2: 30 Probability: | GIGMPLTLLASLAMGVIIGAYQGAWVAFA |
| | 0.678 AA1: | |
| 89007 | 29 AA2: 30 Probability: | MASINASDRRLSSDNLKIIALSAMTLFHA |
| | 0.971 AA1: | |
| 89012 | 33 AA2: 34 Probability: | MYKWYALKRILRGVVMYALLILLFSVLFNSVNE |
| | 0.742 AA1: | MNNAMSSYSFHGIFCAPVRKILFILFFY |
| 89016 | 40 AA2: 41 Probability: | SSFLIPNSSLFA |
| | 0.912 AA1: | MNYKYVLYSIGKLLQILALVLTAPALIAFFETGG |
| 89026 | 57 AA2: 58 Probability: | SFWTRITHTNLTGFLIAIAVSLA |
| | 0.977 AA1: | |
| 89029 | 23 AA2: 24 Probability: | MKNIKDKLILITGGASGIGQLMA |
| | 0.984 AA1: | |
| 89032 | 32 AA2: 33 Probability: | MIRFERSNIALKRLLCVICVIIVSPLFSLSFN |
| | 0.998 AA1: | |
| 89037 | 31 AA2: 32 Probability: | MKLIKKLLGITAIIAAVGFITLPLTGCPAES |
| | 0.997 AA1: | |
| 89049 | 20 AA2: 21 Probability: | MISRAARLILLTAFLAAAHG |
| | 0.735 AA1: | |
| 89050 | 25 AA2: 26 Probability: | MSSIATFSATGLGQGTSVSGGISFA |
| | 0.654 AA1: | |
| 89068 | 26 AA2: 27 Probability: | MRIIWSIVVSFCVLVVISQFTGWVYS |
| | 0.816 AA1: | |
| 89078 | 16 AA2: 17 Probability: | MKIGIIGLGLIGGSFA |
| | 0.873 AA1: | |
| 89090 | 30 AA2: 31 Probability: | MNIETSLFYDLKILLPVILVSLGAIFSLVA |
| | 0.913 AA1: | |
| 89107 | 15 AA2: 16 | PNKSFVVLSAGAAGA |
| 89117 | Probability: | MKIKFKLSIVVITIMAAVVA |

| | | |
|---|---|---|
| | 0.995 AA1: 20 AA2: 21 Probability: | |
| | 0.809 AA1: | |
| 89119 | 37 AA2: 38 | MKSLPFAVQLSLIVFLILIIPFIVLMSYIGSSTLRYS |
| | Probability: | |
| | 0.999 AA1: | |
| 89126 | 34 AA2: 35 | MTMSSVKSLKMFANKRFLVFSCILFFVFALSSIT |
| | Probability: | |
| | 1.000 AA1: | |
| 89138 | 25 AA2: 26 | MKNTIKFRFMGILTLLVTCAIFAVA |
| | Probability: | |
| | 0.999 AA1: | |
| 89152 | 22 AA2: 23 | LKRNILLFAALAALGMTGISCA |
| | Probability: | |
| | 0.999 AA1: | |
| 89160 | 29 AA2: 30 | MNLGLLGAGLVLGIAAVGSGLGIGVAAQA |
| | Probability: | |
| | 0.763 AA1: | |
| 89163 | 18 AA2: 19 | MIKRVCIIGGGNLGMALA |
| | Probability: | |
| | 0.999 AA1: | MKKPKTSLRKMVWGEHKISLRFRFINIALFLLTS |
| 89165 | 43 AA2: 44 | LIMAVVLLS |
| | Probability: | |
| | 0.694 AA1: | MDLSVLIAVGAGIAVFTGLGAGIGIGIATSGFLQAI |
| 89188 | 57 AA2: 58 | SRQPEAAGKMTPMFFVGVALA |
| | Probability: | |
| | 0.875 AA1: | |
| 89194 | 30 AA2: 31 | MAQDILSKQVPFGQVALLVLFSLPSVIALS |
| | Probability: | |
| | 0.737 AA1: | |
| 89206 | 19 AA2: 20 | MLRKSLTHWRWLAAGLVMA |
| | Probability: | |
| | 1.000 AA1: | |
| 89209 | 24 AA2: 25 | MNKTTVYTLKALALLLIAAGLALG |
| | Probability: | |
| | 1.000 AA1: | |
| 89217 | 25 AA2: 26 | MGKALRLVSAAVCAAFLAAPFAVRA |
| | Probability: | |
| | 0.997 AA1: | |
| 89218 | 33 AA2: 34 | MRSFEKKIFVVRQIFAGLCILLMSACTGQKASG |
| | Probability: | |
| | 0.964 AA1: | |
| 89221 | 35 AA2: 36 | VSIRLKTIFILIATVVVITVSTLGISLFVSQYRFN |
| | Probability: | |
| | 0.539 AA1: | |
| 89253 | 29 AA2: 30 | MASINSRAPGLSSDNLKIIALLAMTLYHA |
| | Probability: | |
| | 0.863 AA1: | |
| 89254 | 30 AA2: 31 | MTYKFAFYVATSALLLCVCRADFGPGGGSA |
| 89260 | Probability: | MTIKRSKINKTFQAIKILLLASSILSCTISSGAYK |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 35 AA2: 36 | |
| | Probability: | |
| | 0.994 AA1: | |
| 89276 | 16 AA2: 17 | MKHFLVFLILPALLFA |
| | Probability: | |
| | 0.768 AA1: | MSELGIVTNTNYFLMRRSHAIALYLEKINTALVVLL |
| 89277 | 70 AA2: 71 | AVIALLGLIAVLSKKLRGKALLSAFCFAALAAFA |
| | Probability: | |
| | 1.000 AA1: | |
| 89283 | 35 AA2: 36 | MITYLFRSFKKIVNKTFFIFSCVLFFIIALAAVVA |
| | Probability: | |
| | 0.607 AA1: | |
| 89293 | 34 AA2: 35 | MKKFSTFEGVFLPTILTIFGLVIFMRATFVTGHA |
| | Probability: | |
| | 0.989 AA1: | |
| 89325 | 21 AA2: 22 | MRAKQVLFFLISVLALFAVTS |
| | Probability: | |
| | 1.000 AA1: | |
| 89333 | 22 AA2: 23 | MNKKPHIFIILLLFFISSSLFA |
| | Probability: | |
| | 0.975 AA1: | LFDKLLVNKINLVNITIIKRIRTLSSPFAPLRSLR |
| 89341 | 70 AA2: 71 | PLRPLRENLGIKCALAPARASLLILLVICISSCQA |
| | Probability: | |
| | 0.849 AA1: | MLATIQRMISWIPSLLNGAKITISLTILAVSAGLVL |
| 89345 | 43 AA2: 44 | SLFIALG |
| | Probability: | |
| | 0.999 AA1: | |
| 89359 | 21 AA2: 22 | LRKKFSYIVLAAALIMPAVVS |
| | Probability: | |
| | 0.998 AA1: | |
| 89402 | 25 AA2: 26 | MKRLDWLFMFSLFALIGIGLALVYS |
| | Probability: | |
| | 0.995 AA1: | |
| 89436 | 20 AA2: 21 | MRKFLIACSLGAVATFGLFS |
| | Probability: | |
| | 0.641 AA1: | |
| 89441 | 20 AA2: 21 | AYLWRQAVGGAVIGRPTVQA |
| | Probability: | |
| | 0.851 AA1: | |
| 89451 | 23 AA2: 24 | ANWNAATGLSAGTLTIDAAIPAA |
| | Probability: | |
| | 0.975 AA1: | MSDKVIHTSDTKTLIKNNIRNYSMFIMLALIML |
| 89460 | 41 AA2: 42 | IFAPLTNG |
| | Probability: | |
| | 0.943 AA1: | |
| 89469 | 33 AA2: 34 | MNIAEYSVKHPVTIIMIFALVIGMAVTMLPTLA |
| | Probability: | |
| | 0.622 AA1: | |
| 89475 | 29 AA2: 30 | MDLVKAGLIIAAAHLVVMTAVLLLQYAAA |
| 89497 | Probability: | VTKKRLIIGGGAVVLLLAFIFSLPRPLFN |

| | | |
|---|---|---|
| | 0.997 AA1: 29 AA2: 30 Probability: | |
| | 0.998 AA1: | |
| 89507 | 21 AA2: 22 | MMKKLFFTFFLIFTLFLPLSA |
| | Probability: 0.910 AA1: | |
| 89513 | 23 AA2: 24 | MKKSILILFTLAVFLFGISGCKK |
| | Probability: 0.999 AA1: | |
| 89519 | 29 AA2: 30 | MCIIRNKLAKFAQVAALVLALTFTFSCSG |
| | Probability: 0.917 AA1: | |
| 89523 | 41 AA2: 42 | MSGKIIHTTDTKTLIKSNIRNYSMFIMLALIMLIFAPLTKG |
| | Probability: 0.821 AA1: | MDKKANNMNNFSSGILPPVLVTGLFFGLIYMQNNFSTAV |
| 89574 | 71 AA2: 72 | FLVFNAMVIFFLAGMRFRYFFAAIAMILPVSA |
| | Probability: 0.931 AA1: | |
| 89588 | 32 AA2: 33 | MKDNILKFLKYAILIAASFIVLLPFVSVLITA |
| | Probability: 0.747 AA1: | |
| 89617 | 24 AA2: 25 | MIRSIIWTIVFSIVAALLQSTVLE |
| | Probability: 0.993 AA1: | |
| 89644 | 26 AA2: 27 | MASSIRSKILIPCVGIAVVCVVASTA |
| | Probability: 0.525 AA1: | MYGGGRGRNNKGGGAAVIGLILMILAPIAAML |
| 89645 | 44 AA2: 45 | IQMAISRSREYA |
| | Probability: 1.000 AA1: | |
| 89647 | 29 AA2: 30 | MNSLKSIKPILIVATAVVLLTASLFLSMA |
| | Probability: 0.509 AA1: | |
| 89648 | 21 AA2: 22 | ETTYAVTVVSGAGGGSYPAGA |
| | Probability: 0.988 AA1: | |
| 89661 | 21 AA2: 22 | MKIKIIGLMVIIGFFITSCEA |
| | Probability: 0.569 AA1: | |
| 89662 | 41 AA2: 42 | MNNMSQSLYQLKKNTGSFVFKFCRAILLFGLCFLIVQPLLN |
| | Probability: 0.904 AA1: | |
| 89672 | 21 AA2: 22 | AKNKILALYAANAPFGANVVG |
| | Probability: 0.996 AA1: | |
| 89677 | 27 AA2: 28 | MVSRGFKKIILFLIGCMFFPLSLYSLG |
| | Probability: 0.974 AA1: | |
| 89678 | 25 AA2: 26 | MKKISTLLIGICLLLTVIGCQNPFQ |
| 89682 | Probability: | MKKKNKCLLILLFLFAMPLFS |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 0.891 AA1: | |
| 89697 | 25 AA2: 26 | LIGNTIVNVVTSALATMLFMGLFGA |
| | Probability: | |
| | 0.942 AA1: | |
| 89702 | 30 AA2: 31 | MIKAKSVHITVFIKIFSILLFFSAITPLFG |
| | Probability: | |
| | 0.994 AA1: | |
| 89703 | 24 AA2: 25 | IMIRLCKLLLIILFTIPAVFTLSA |
| | Probability: | |
| | 0.996 AA1: | |
| 89720 | 32 AA2: 33 | LRIIKGDASILLLVVIFLLLAFGVVFSVMAFR |
| | Probability: | |
| | 0.574 AA1: | |
| 89753 | 22 AA2: 23 | MNSIASQTNLLSMNAAIEAAHA |
| | Probability: | |
| | 0.536 AA1: | |
| 89760 | 21 AA2: 22 | GGGLLSQAGGMIAAACGGESA |
| | Probability: | |
| | 0.996 AA1: | MKRTLFLVGAVIVVLLLLFPFPPGIMIGGIHMTRH |
| 89775 | 54 AA2: 55 | LLFIFGLVMLIAAAVIHA |
| | Probability: | |
| | 0.915 AA1: | |
| 89776 | 33 AA2: 34 | ATSNGRAIYFNSTGTLTISGGVVSANTAGYAVA |
| | Probability: | |
| | 0.999 AA1: | |
| 89802 | 20 AA2: 21 | MKKFLYFAVILALAILTVFA |
| | Probability: | |
| | 0.921 AA1: | |
| 89812 | 23 AA2: 24 | MNKRQTTIITALSIIAFFLAFMV |
| | Probability: | |
| | 0.579 AA1: | |
| 89840 | 25 AA2: 26 | AKAWLTAALIVIVVLSGCKTTDKAA |
| | Probability: | |
| | 0.832 AA1: | |
| 89849 | 30 AA2: 31 | SIRSIKSARPIRAERQKELTVIALRALASA |
| | Probability: | |
| | 0.999 AA1: | |
| 89878 | 32 AA2: 33 | PTTPAAPSAPTIPAALSAPTTPAAPSASNAQA |
| | Probability: | |
| | 0.563 AA1: | VKKFAIHNRKNWTAFSWKESTITPLLNEVIRL |
| 89881 | 51 AA2: 52 | QGKICGLMLAMGFSQKTEA |
| | Probability: | |
| | 1.000 AA1: | |
| 89894 | 31 AA2: 32 | MTKRLHPSSITFVPLCLCVTLFAAPFLAAQS |
| | Probability: | |
| | 1.000 AA1: | |
| 89902 | 24 AA2: 25 | MKAKLCSILVIFLSLSLITCNVFN |
| 89916 | Probability: | LSLSKAGASRLLLGLIILLIVLGAAA |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| | 0.996 AA1: | |
| 89930 | 24 AA2: 25 | MQFILILLLAAPLLAAVSGESRYA |
| | Probability: | |
| | 0.998 AA1: | |
| 89954 | 30 AA2: 31 | MKALVSWLKRKKSAFCILTGIISVSLFVIS |
| | Probability: | |
| | 0.945 AA1: | |
| 89960 | 28 AA2: 29 | LSGLMMPIMMFIGNFGYVAVCVAGAALA |
| | Probability: | |
| | 0.996 AA1: | |
| 89961 | 25 AA2: 26 | MKWLAVIFAIFATFATFGIGNMTQA |
| | Probability: | |
| | 1.000 AA1: | |
| 89965 | 26 AA2: 27 | MILFNRQARTLLAVLFLWAAVSTALA |
| | Probability: | |
| | 0.994 AA1: | |
| 89981 | 26 AA2: 27 | MKKVPFFRLFFVGLAVILLLISGCKS |
| | Probability: | |
| | 1.000 AA1: | |
| 89989 | 27 AA2: 28 | MKKTKIDASAFLLILIVLLLAGGAVFA |
| | Probability: | |
| | 1.000 AA1: | |
| 90027 | 23 AA2: 24 | MKAKHFLVFVLALIATFTLGCAA |
| | Probability: | |
| | 1.000 AA1: | |
| 90028 | 24 AA2: 25 | MNSKFSGLTAVIFFMACVPFSVKA |
| | Probability: | |
| | 0.682 AA1: | |
| 90030 | 26 AA2: 27 | DADFSAQKKRRGIFSVALFSGNVSLS |
| | Probability: | |
| | 1.000 AA1: | |
| 90044 | 31 AA2: 32 | LQSSSISKFTALLLYVASILTGAAAVRPASA |
| | Probability: | |
| | 1.000 AA1: | |
| 90054 | 22 AA2: 23 | MRCFFPVGLFCFLFVFSFSIYA |
| | Probability: | |
| | 0.973 AA1: | |
| 90072 | 42 AA2: 43 | MNKKVSEIFSRRPLYIQILFTSFAFLAMVILSYIFTSRIARA |
| | Probability: | |
| | 0.992 AA1: | |
| 90084 | 30 AA2: 31 | MAVNIKPVQPVIAFITLFVTLLCFSCASSA |
| | Probability: | |
| | 1.000 AA1: | |
| 90085 | 29 AA2: 30 | LKFFGHRGFKLCLFALLLAGLILAAPLYG |
| | Probability: | |
| | 0.736 AA1: | MSVILSKLLVTLLNPVGLTLVFLFCGLLFMLIWR |
| 90116 | 59 AA2: 60 | KTKLSAIFFIGGFAVLLASSLPVTA |
| 90121 | Probability: | MGLYFRISFLFTIILTSTANG |

| | | |
|---|---|---|
| | 0.857 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 0.954 AA1: | |
| 90126 | 26 AA2: 27 | MGKDLKILKYLLCLLLVLPGLGRLEA |
| | Probability: | |
| | 1.000 AA1: | |
| 90150 | 23 AA2: 24 | MKTKLLKLCFLLLIVFSVSPAFS |
| | Probability: | |
| | 0.530 AA1: | |
| 90152 | 17 AA2: 18 | MLVFPLLCALFIGCAKS |
| | Probability: | |
| | 1.000 AA1: | |
| 90153 | 23 AA2: 24 | MYALLKNLLLLLCAVLCAGSVFS |
| | Probability: | |
| | 1.000 AA1: | |
| 90157 | 26 AA2: 27 | MNRTGRTAFLAAIIALLAGAAAGAWA |
| | Probability: | |
| | 1.000 AA1: | MRGFRHSSGMNKTVRTVVWSVMAAVLA |
| 90158 | 39 AA2: 40 | VAGSGWLFGASA |
| | Probability: | |
| | 1.000 AA1: | |
| 90163 | 23 AA2: 24 | MKNTLWVIAILTAFTLSTTSCFA |
| | Probability: | |
| | 1.000 AA1: | |
| 90165 | 29 AA2: 30 | MDYRSYFAAFLAVFLAVLSTFLSTTTAQA |
| | Probability: | |
| | 0.982 AA1: | |
| 90168 | 26 AA2: 27 | MEKIKLSLGILAVTGLVSCASGPRIS |
| | Probability: | |
| | 0.998 AA1: | MKKISASVRKGANAYLKRQYTGVAMFFVVMFVVLGIM |
| 90171 | 74 AA2: 75 | AINGLLTPFVPFAFITGGFFSGLSGFIGMKIATSANA |
| | Probability: | |
| | 1.000 AA1: | |
| 90181 | 32 AA2: 33 | MGNRFLGRKIKKAFKIAGIVFIAVILMAVVFA |
| | Probability: | |
| | 0.922 AA1: | |
| 90182 | 19 AA2: 20 | VSIALDLVSFAQAAPPALA |
| | Probability: | |
| | 0.984 AA1: | |
| 90189 | 20 AA2: 21 | MKKVFVLFVVALCAVSLVLS |
| | Probability: | |
| | 0.919 AA1: | |
| 90192 | 22 AA2: 23 | MINRRRSVSIYFLIPAILLVAA |
| | Probability: | |
| | 1.000 AA1: | |
| 90214 | 22 AA2: 23 | MKKLFVLIALVYLFLPLSFSSA |
| | Probability: | |
| | 0.914 AA1: | |
| 90246 | 26 AA2: 27 | MRLAGKTALISGAAQGLGAAMARRFA |
| 90249 | Probability: | MRKLVKYLKPFTLLFILAVLILFLQVFCDLS |

| | | |
|---|---|---|
| | 0.940 AA1: 31 AA2: 32 Probability: | |
| | 0.972 AA1: | |
| 90253 | 33 AA2: 34 | MNGNNRKKIVPTVFLVAGIMMMFIDSGAAKVLG |
| | Probability: | |
| | 0.793 AA1: | |
| 90274 | 25 AA2: 26 | MDKRVFGADLLLLLTAAVWGFGFVA |
| | Probability: | |
| | 0.999 AA1: | |
| 90275 | 19 AA2: 20 | KKSFIFLFLLLFIHSFAFS |
| | Probability: | |
| | 0.632 AA1: | |
| 90305 | 36 AA2: 37 | MSQNPTEVYWNSDRKMRFLFAASGLAAFFALLYYLS |
| | Probability: | |
| | 0.999 AA1: | |
| 90307 | 27 AA2: 28 | MYSKGKFFSIAAILSIIIMLTILIVTA |
| | Probability: | |
| | 1.000 AA1: | |
| 90323 | 20 AA2: 21 | MRKLILTTAAVAAAIAGAFA |
| | Probability: | |
| | 0.838 AA1: | |
| 90326 | 21 AA2: 22 | VNSITVCMGSSCFGKGNSANA |
| | Probability: | |
| | 0.973 AA1: | |
| 90331 | 32 AA2: 33 | MKKALQYIILTCAVSWAAAGVAIWLGMRTAQG |
| | Probability: | |
| | 0.510 AA1: | MLTNNIYTPISTFFRKLRQRWVLFLDIRDDNYTYR |
| 90336 | 58 AA2: 59 | VITLVSAVFLAVFAIVLGLSLLG |
| | Probability: | |
| | 1.000 AA1: | |
| 90341 | 27 AA2: 28 | MRKTKIDASAFLLILIVLLLTGGVVFA |
| | Probability: | |
| | 0.952 AA1: | |
| 90342 | 22 AA2: 23 | MGNFSNKAFLLSLIFLFFLGCA |
| | Probability: | |
| | 0.997 AA1: | |
| 90343 | 32 AA2: 33 | MNKKLISINKIICLFFVALCFLFITCNGPDNN |
| | Probability: | |
| | 0.798 AA1: | |
| 90350 | 26 AA2: 27 | MYTFNAYAVAVVALVILVFTTMMFLA |
| | Probability: | |
| | 0.890 AA1: | |
| 90353 | 37 AA2: 38 | LRSFLEKFKILNSFRAQIIIVLAVILIISFSTFLFFS |
| | Probability: | |
| | 1.000 AA1: | |
| 90354 | 35 AA2: 36 | VRILKYVTRKRLIIAGALLVLLLAFIFSLPRPLFN |
| | Probability: | |
| | 1.000 AA1: | |
| 90356 | 28 AA2: 29 | MYITNRLVTMAVMLFLLGLAVFALMALA |
| 90360 | Probability: | MDTITVCMGSSCFGKGNSANA |

| | | |
|---|---|---|
| | 0.546 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 0.724 AA1: | |
| 90369 | 21 AA2: 22 | MKELQYDAVVIGAGAAGMAAA |
| | Probability: | |
| | 0.934 AA1: | MRYRLRYKLAFCTISLVLLVDLLVSAPLFFMLIR |
| 90377 | 53 AA2: 54 | GHQEAPLKAMFIISAAALA |
| | Probability: | |
| | 0.633 AA1: | |
| 90382 | 20 AA2: 21 | MKRLYIFPAWVLIFSSTIFS |
| | Probability: | |
| | 0.972 AA1: | |
| 90430 | 22 AA2: 23 | MKIYGYMTALLSVLMMTVYCAN |
| | Probability: | |
| | 1.000 AA1: | MKDNGFPDKNIPSRRKIPFSWIWFASTILLCAV |
| 90431 | 44 AA2: 45 | FAFASVPSVLA |
| | Probability: | |
| | 1.000 AA1: | |
| 90446 | 23 AA2: 24 | MTKIRASRLILAAVMAVSVGAFS |
| | Probability: | |
| | 1.000 AA1: | |
| 90453 | 21 AA2: 22 | MKKSIFIALCVIMLAFQSLFA |
| | Probability: | |
| | 0.545 AA1: | MKYKKTTVKKGASIGANATIVCGVTIGEYAL |
| 90478 | 39 AA2: 40 | IGAGSVVT |
| | Probability: | |
| | 0.933 AA1: | |
| 90480 | 22 AA2: 23 | MKTNYIKIPFFWLIFTAGLGFA |
| | Probability: | |
| | 0.996 AA1: | |
| 90483 | 22 AA2: 23 | MQSGKILSAVTLLLFSLVSVYA |
| | Probability: | |
| | 0.998 AA1: | |
| 90519 | 30 AA2: 31 | MKQAKNIKSSAAAARVRAVLAAFMLLSAAS |
| | Probability: | |
| | 0.939 AA1: | |
| 90535 | 21 AA2: 22 | DRNLLALSAHNPALCSQLSLA |
| | Probability: | |
| | 0.996 AA1: | |
| 90539 | 22 AA2: 23 | VPYILSLISAFLLLSCASNVHR |
| | Probability: | |
| | 0.986 AA1: | MPRIFPTFRFPRLPRLIRPHRGFFFLCGICLALVL |
| 90541 | 44 AA2: 45 | FFVINAVFL |
| | Probability: | |
| | 0.961 AA1: | |
| 90542 | 21 AA2: 22 | MVNMRLSIVLIICLLILPACA |
| | Probability: | |
| | 0.993 AA1: | |
| 90546 | 22 AA2: 23 | LRIVLVGFGLLAGSVASAIRQA |
| 90554 | Probability: | MKRAIFAVIVLAFAAACTAPSDMA |

| | | |
|---|---|---|
| | 1.000 AA1: 24 AA2: 25 Probability: 0.963 AA1: | |
| 90556 | 17 AA2: 18 Probability: 0.995 AA1: | MTIREALAGGSAALAAA |
| 90560 | 28 AA2: 29 Probability: 0.715 AA1: | MGGVNDKRAKQMTKKTAACLLAFPPVLA |
| 90568 | 33 AA2: 34 Probability: 0.992 AA1: | MEKRKVNRAAVFYPCTMGVFILLIIILSFTAGA |
| 90582 | 24 AA2: 25 Probability: 1.000 AA1: | MTCKRIAPAILAALLPLSLAVLMG<br>MLNTKKDPVLIVKICIIILIGAFAFSFIGSFF |
| 90602 | 38 AA2: 39 Probability: 1.000 AA1: | AQRAQA |
| 90603 | 38 AA2: 39 Probability: 1.000 AA1: | MLNQKKDPILIVKVCLLILIGAFAFSFISSLFNARAMA |
| 90620 | 36 AA2: 37 Probability: 0.960 AA1: | MNNIKMGPKLIASFLFLAALAAFMGIRSISSLKSVA |
| 90641 | 24 AA2: 25 Probability: 1.000 AA1: | MAFVAKRLVITLITLFLVSVITFA |
| 90643 | 23 AA2: 24 Probability: 0.937 AA1: | MRRKSTFKSAAAVIFAALFFAMA |
| 90647 | 26 AA2: 27 Probability: 0.503 AA1: | MNNMKGNKLPVLLVIVLMSACNIFLG |
| 90649 | 27 AA2: 28 Probability: 0.831 AA1: | MEYLIKAILCISGVFIIGIGVSFMRYA |
| 90653 | 23 AA2: 24 Probability: 0.999 AA1: | MAKLGSIIQVVGPVVDVSFSASA |
| 90667 | 23 AA2: 24 Probability: 0.936 AA1: | MLGRIILSIAIGVSAVSFLSAQA |
| 90679 | 28 AA2: 29 Probability: 0.736 AA1: | MRRITAADIAITAAFWILTALSSAHCAA |
| 90681 | 26 AA2: 27 Probability: 0.983 AA1: | MNLAPIIGIVGALAMVIYGMLGSGNA |
| 90695 | 28 AA2: 29 | LRVSRKILAGRLGIMTAGTDFSASVAWV |
| 90713 | Probability: | VASRVCLIAAALFAAILCRSAPA |

| | | |
|---|---|---|
| | 0.994 AA1: 23 AA2: 24 Probability: | |
| | 0.969 AA1: | |
| 90728 | 37 AA2: 38 | MSKEQITKLRYASLSLLFVTCYLLFVLPSLPASPLHS |
| | Probability: | |
| | 0.957 AA1: | |
| 90729 | 32 AA2: 33 | MRDVIAMRNFVIVMASIAAAAAVVILFLVLNS |
| | Probability: | |
| | 0.922 AA1: | |
| 90734 | 21 AA2: 22 | KKDALMGLIKAAVGAAGGVLA |
| | Probability: | |
| | 0.836 AA1: | |
| 90742 | 32 AA2: 33 | MTELAKQRLAKFRHNKLAWFSFVILAVAYLLS |
| | Probability: | |
| | 0.941 AA1: | |
| 90750 | 26 AA2: 27 | MRLAGKTALISGAAQGLGAAMACRFA |
| | Probability: | |
| | 1.000 AA1: | |
| 90767 | 20 AA2: 21 | MKKFLLLFVAGYFLSVTVEA |
| | Probability: | |
| | 0.519 AA1: | |
| 90776 | 20 AA2: 21 | MPKITFLGAGSTVFARNVLG |
| | Probability: | |
| | 0.981 AA1: | MIIASGTVISSVFSAKIIKRLGTGIVTSVSVLMTAV |
| 90781 | 66 AA2: 67 | ALAGFSLSKSFYFLCFFAIPLGLGAGSVDA |
| | Probability: | |
| | 1.000 AA1: | |
| 90825 | 21 AA2: 22 | KNIIGSLCLSLLFVVIVPAFA |
| | Probability: | |
| | 0.935 AA1: | MPNYLSRNPMLYPVAGFLALISLGTILLL |
| 90829 | 55 AA2: 56 | MPFSSKDGLSFSDALFMATSASCVNG |
| | Probability: | |
| | 0.996 AA1: | |
| 90831 | 24 AA2: 25 | VSTFKKTALVLIISSLLFAGMVYL |
| | Probability: | |
| | 0.922 AA1: | |
| 90848 | 21 AA2: 22 | MPTIVVVGIVFLAVIALIIFA |
| | Probability: | |
| | 1.000 AA1: | |
| 90851 | 29 AA2: 30 | MTRFSFNRLGRALSVASAAVVGAVFSVGA |
| | Probability: | |
| | 0.989 AA1: | MEFLKDLWNFMRQRKKFWLLPVIVILVLFGFL |
| 90858 | 42 AA2: 43 | MVFTSGTAIA |
| | Probability: | |
| | 0.955 AA1: | |
| 90868 | 22 AA2: 23 | VTKKRLIIGGAAVFLLLAFIFS |
| | Probability: | |
| | 0.707 AA1: | MDLQVLIAVGAGIAVLTGAGAGIGIGIATSGFLQAISR |
| 90874 | 57 AA2: 58 | QPEASGKMTPMFFVGVALA |
| 90881 | Probability: | MNLRPLAIILIFLFLPAIIFA |

| | | |
|---|---|---|
| | 0.987 AA1: 21 AA2: 22 Probability: | |
| | 0.997 AA1: | |
| 90898 | 29 AA2: 30 | MKKLKKTSFARFFTTVSLLMILSITVSLS |
| | Probability: 0.999 AA1: | |
| 90899 | 29 AA2: 30 | MKKPRKTSFAMLFITVSLLMILSITVSLS |
| | Probability: 1.000 AA1: | |
| 90915 | 20 AA2: 21 | VVVILSLALMLTVSSLAAYA |
| | Probability: 0.604 AA1: | MKLESNIPGGPIAEKWTRHKFELKLVNPAN |
| 90916 | 49 AA2: 50 | KRKFTVIVVGTGLAGASAA |
| | Probability: 0.977 AA1: | |
| 90919 | 23 AA2: 24 | MKNIKDKLILITGGASGIGQLMA |
| | Probability: 0.998 AA1: | |
| 90937 | 35 AA2: 36 | LGIFIFYIPSNMTFSPKKLALAAALGIAITLTLTA |
| | Probability: 1.000 AA1: | |
| 90942 | 25 AA2: 26 | MFKSGVTIMIPLLFAAVGGLFPALA |
| | Probability: 0.999 AA1: | |
| 90943 | 21 AA2: 22 | APIRVALALFLFLYLGPAIFA |
| | Probability: 0.715 AA1: | |
| 90958 | 22 AA2: 23 | EGTRKIASIPGTLSDFIFSAFV |
| | Probability: 0.576 AA1: | |
| 90959 | 27 AA2: 28 | MRRNNDKIDTRTAAKAVSALRSFFRFA |
| | Probability: 0.920 AA1: | |
| 90968 | 31 AA2: 32 | MNFYKINRLLHNKSGFQLLMPLLILFLLVIA |
| | Probability: 0.961 AA1: | |
| 90994 | 35 AA2: 36 | MGVYMRLFTKPRLTLLLATTIALFAICLSGCGGSG |
| | Probability: 1.000 AA1: | |
| 91010 | 22 AA2: 23 | MKLNAKLLVFGACFLMARPLFA |
| | Probability: 0.973 AA1: | |
| 91011 | 36 AA2: 37 | MMSLLRITKAVFKKYAQVFAVFLTFALMVLFSYSFT |
| | Probability: 0.992 AA1: | |
| 91012 | 22 AA2: 23 | NRLLFLFSLSFVLAVSSCSRLG |
| | Probability: 0.996 AA1: | |
| 91016 | 28 AA2: 29 | MKNKYTKLFTPLFALILFSSCMSINRNR |
| 91034 | Probability: | MQIFQKSNLNERRTSIKRKFIVFSSVLFLA |

| | | |
|---|---|---|
| | 1.000 AA1: 52 AA2: 53 Probability: | IVLLGSGAFLLSMRQIVRNSTA |
| | 0.976 AA1: | |
| 91035 | 22 AA2: 23 Probability: | LKRTFALFTALTVLGMIGISCS |
| | 0.998 AA1: | |
| 91042 | 23 AA2: 24 Probability: | MSNIGKITAFVIFIFVLLHPVFA |
| | 0.732 AA1: | |
| 91047 | 28 AA2: 29 Probability: | MYPDTPGIRALFFLVVVGTLTLCGGLTA |
| | 0.793 AA1: | |
| 91055 | 22 AA2: 23 Probability: | GNYNRYAAATMFSLFFGGKIFA |
| | 0.578 AA1: | |
| 91058 | 24 AA2: 25 Probability: | TVITTNTGLAVSAINGVVINNVNA |
| | 0.908 AA1: | |
| 91098 | 23 AA2: 24 Probability: | VTLERPVLAIAALLIFPLAIVLA |
| | 0.997 AA1: | |
| 91119 | 37 AA2: 38 Probability: | IVQKINNTWTIVKSSHPIFTAVLMILFCAMLLVLTLA |
| | 0.989 AA1: | |
| 91136 | 26 AA2: 27 Probability: | MKKSKYISFFIIFFIFAGGLCLSCKA |
| | 1.000 AA1: | |
| 91138 | 22 AA2: 23 Probability: | MRCFFPVGLFCFLFIFSFSIYA |
| | 0.540 AA1: | |
| 91186 | 13 AA2: 14 Probability: | MALTLSALGIVLA |
| | 0.991 AA1: | |
| 91187 | 36 AA2: 37 Probability: | MMSLLRITKTVFKKYAQVFAVFLAFALMVIFSYSFT |
| | 0.622 AA1: | |
| 91198 | 18 AA2: 19 Probability: | MNRKIFTIIFLIIISACG |
| | 0.953 AA1: | |
| 91207 | 20 AA2: 21 Probability: | MPKFVLLLLVGLFVQIPIFS |
| | 0.861 AA1: | |
| 91212 | 41 AA2: 42 Probability: | LRIFNLCGIYIILALSLNLINGFAGLFSLGHAGFIALGAYT |
| | 0.695 AA1: | |
| 91213 | 25 AA2: 26 Probability: | MIPGLKCFFAAVMGGIGNISGAVLG |
| | 0.812 AA1: | |
| 91220 | 18 AA2: 19 | MTGAALAAGLLRGAFSLA |
| 91232 | Probability: | MKKLFGTLCALIIALSILCLASCTNNNA |

| | | |
|---|---|---|
| | 1.000 AA1: 28 AA2: 29 Probability: 0.748 AA1: | MRTATFLTVGRSLRWPVFLLTAGVLFLGG |
| 91249 | 46 AA2: 47 Probability: 0.673 AA1: | IMISGCSPDSDGNPAGA |
| 91250 | 18 AA2: 19 Probability: 1.000 AA1: | MPAILNLIWLLPILAVFA |
| 91259 | 26 AA2: 27 Probability: 1.000 AA1: | MKKTFFLTAAAVMLTAAMTGCTNVQN |
| 91262 | 21 AA2: 22 Probability: 0.578 AA1: | LKKQIAVVLCMMALVFQPLFA |
| 91265 | 22 AA2: 23 Probability: 0.969 AA1: | MSVSCKYWVLGVYCLVSFPLFS |
| 91311 | 28 AA2: 29 Probability: 1.000 AA1: | MPKYKNLAKTKCLLFLTVIIIAVTGCDA |
| 91319 | 23 AA2: 24 Probability: 0.708 AA1: | MTRLNKLLFFIFFSLLIVLSAAA |
| 91322 | 22 AA2: 23 Probability: 0.937 AA1: | MTMNAIFMMAIIPLVSLGITMM |
| 91339 | 39 AA2: 40 Probability: 0.999 AA1: | MFYTVSSSGGFQGTPDQIIFIVFVCLLFFFICKSAPAVA |
| 91352 | 21 AA2: 22 Probability: 0.596 AA1: | MKKIITTIIVILVIIIGIAVA |
| 91355 | 34 AA2: 35 Probability: 0.677 AA1: | MHVTNDKEQKTKKTISRYMQILFVAIAFGLMVFA |
| 91392 | 23 AA2: 24 Probability: 0.636 AA1: | MLPIICFISITLLFLDFTGTVHT |
| 91405 | 26 AA2: 27 Probability: 0.843 AA1: | MVDNAVLQNILLFGPVFLLIAARALA |
| 91406 | 21 AA2: 22 Probability: 1.000 AA1: | LISALVIGLVVAILQATTSIQ |
| 91420 | 47 AA2: 48 Probability: 0.511 AA1: | MNIRSKIMLSIAVGAAVATLAVCTMSVFSVKSIVVVAAIISVFCAVA |
| 91423 | 19 AA2: 20 | MTIETIFKLVAIGILASVA |
| 91428 | Probability: | MKNNKVKLIPLLGLLSVILYFLHVILG |

| | | |
|---|---|---|
| | 0.813 AA1: 27 AA2: 28 Probability: 0.979 AA1: | |
| 91441 | 19 AA2: 20 Probability: 1.000 AA1: | MKRLFILINLLFIPILCFA |
| 91448 | 25 AA2: 26 Probability: 1.000 AA1: | MKTNRRFLLAAIFGLAITFTLSCSG |
| 91472 | 25 AA2: 26 Probability: 0.997 AA1: | MRRLRILICLIYVLSFFPVAAVAQA |
| 91475 | 29 AA2: 30 Probability: 0.999 AA1: | MKIKFKLSIIVIAIMAVVVSGIATLLLWQ |
| 91476 | 23 AA2: 24 Probability: 0.992 AA1: | MKYVSKALFLLILFTGMAGIGYA |
| 91515 | 22 AA2: 23 Probability: 0.936 AA1: | MKTLNRYGWPLLSAALLFAVVA |
| 91543 | 29 AA2: 30 Probability: 0.965 AA1: | ISDNLLLIPLVVIVVLAMLFFSFRRFTFV |
| 91549 | 27 AA2: 28 Probability: 0.903 AA1: | MRKEPLFASGVAALTLAFFLIGGPLHS |
| 91552 | 39 AA2: 40 Probability: 0.999 AA1: | VKVKSGRGKIPGFGLSLGLSLSYISFLILIPISALIIYA |
| 91562 | 21 AA2: 22 Probability: 1.000 AA1: | MKSITKLTATLALAITLTFTA |
| 91574 | 17 AA2: 18 Probability: 0.998 AA1: | MTKKNAAFLLALVPALA |
| 91589 | 22 AA2: 23 Probability: 1.000 AA1: | MSIKNTVIIGLTIMLLAVSCSA |
| 91595 | 30 AA2: 31 Probability: 1.000 AA1: | MNMKLIFLLRKNIFTALFVFIVGVSLYAQS |
| 91597 | 31 AA2: 32 Probability: 0.983 AA1: | MKIKTRILIQTLSASLALILVLGLVFFISVA<br>KQKIFDIDKNIDKNINKNINKNINKNINKNKKINK |
| 91601 | 61 AA2: 62 Probability: 0.914 AA1: | IQNKTERGTSMRLKSVITICGMLTLA |
| 91611 | 18 AA2: 19 | MNNKKYTFLALIIGTGLA |
| 91626 | Probability: | MKIGIKLTALMTALGLFTIASIS |

| | | |
|---|---|---|
| | 0.991 AA1: 23 AA2: 24 Probability: | |
| | 0.998 AA1: | |
| 91628 | 23 AA2: 24 | MKLKPSITAIFIFFFIALTAVYS |
| | Probability: 0.576 AA1: | |
| 91689 | 18 AA2: 19 | MNVGVAGLGLIGGSLAKA |
| | Probability: 1.000 AA1: | |
| 91704 | 31 AA2: 32 | MRRIKNNIIFRLWALLLGFFPVIMLPSLVMA |
| | Probability: 0.875 AA1: | |
| 91723 | 23 AA2: 24 | MSDALLSPAVGVSMYAVSTAACA |
| | Probability: 0.992 AA1: | |
| 91730 | 20 AA2: 21 | LVEKGICIAALVSLVLIAVA |
| | Probability: 1.000 AA1: | |
| 91742 | 19 AA2: 20 | MKNKVSLFLLCAIASGLWA |
| | Probability: 0.826 AA1: | DAGSGSSSSSVAKSSSSSSVRSSSSS |
| 91746 | 50 AA2: 51 | SLVQSSSSSVVSSSSIEQSSSSSS |
| | Probability: 1.000 AA1: | |
| 91755 | 22 AA2: 23 | MLKVNRIVCIFLLVLSFSFLFA |
| | Probability: 0.999 AA1: | |
| 91761 | 39 AA2: 40 | MPQKEKGVMKKGFGLRFSVVVMAAVIGVLFCGSALPAFA |
| | Probability: 0.961 AA1: | |
| 91798 | 17 AA2: 18 | MDKNALALVALSVRSLA |
| | Probability: 0.775 AA1: | |
| 91804 | 37 AA2: 38 | EGQNARFSFSPPKTYVRYAVLGLFVVAIFFGISAVVS |
| | Probability: 0.999 AA1: | |
| 91810 | 34 AA2: 35 | MNLRIKLTVIILAMILVSVAVLSVFNLTRSSALQ |
| | Probability: 0.988 AA1: | VNNFRKRVYGGRANTVRPYAAGGGVDTVWKMLSR |
| 91832 | 54 AA2: 55 | TMLIVAAFVCAFGFSSAASA |
| | Probability: 0.731 AA1: | |
| 91836 | 20 AA2: 21 | MKKIIIGLGSCGIASGGLKV |
| | Probability: 0.998 AA1: | |
| 91841 | 33 AA2: 34 | MNKITRFLIASVLVTALIFTACTKKTGTTVSVG |
| | Probability: 0.932 AA1: | |
| 91850 | 28 AA2: 29 | VIIPRRITDIIFAWLFRAFSFTGAAALG |
| 91870 | Probability: | MSQQPQKHALVTGATSGIGLAIA |

| | | |
|---|---|---|
| | 0.794 AA1: 23 AA2: 24 Probability: 0.901 AA1: | |
| 91871 | 25 AA2: 26 Probability: 0.740 AA1: | MNAIRLISHLLISLMLFTSCGKNNS |
| 91872 | 35 AA2: 36 Probability: 0.790 AA1: | MWRDRVREITNKLKYHKGPVITVAALAVLAIVYFA |
| 91873 | 25 AA2: 26 Probability: 0.803 AA1: | MYHKARVITVAALAVLVIVYFTVGG |
| 91925 | 23 AA2: 24 Probability: 1.000 AA1: | ALFIACAAAIGGFLFGFDTSVIN |
| 91940 | 27 AA2: 28 Probability: 0.998 AA1: | MKNAKLILLAASFAFAASLTISCSSDS |
| 91941 | 25 AA2: 26 Probability: 0.998 AA1: | MKSILKRSFVFVAVAVLIVSAYFIA |
| 91959 | 32 AA2: 33 Probability: 0.639 AA1: | MSNRARIFTGAVVSLFIIAAAVSLGSSGISLR |
| 91962 | 20 AA2: 21 Probability: 0.995 AA1: | FLVILLCLSGAAYSINLFWA |
| 91971 | 26 AA2: 27 Probability: 0.957 AA1: | MKSKFFSKTVAWLNLLLAALFLAISG |
| 91973 | 18 AA2: 19 Probability: 0.541 AA1: | TGLSLATSGVISGTPSTA |
| 91988 | 30 AA2: 31 Probability: 0.998 AA1: | LSIASTSVRRPVLTLVLSAFIVLFGLIGLK |
| 91999 | 25 AA2: 26 Probability: 0.878 AA1: | MVKNMNIFKIFFFSVLILLTSSVFS |
| 92000 | 27 AA2: 28 Probability: 0.993 AA1: | MDKNPMERTLLVCCGTGCIANGAKAVA |
| 92003 | 20 AA2: 21 Probability: 0.999 AA1: | MRALFFTLILAIIINVGLFA |
| 92028 | 24 AA2: 25 Probability: 0.782 AA1: | MKYFPLVQVFLFLFGLFLSVPAVS |
| 92029 | 16 AA2: 17 | DRVKIGCMTGGLGTLA |
| 92040 | Probability: | MKNTLNVLLVRVLGIIALVAVIGFSMTACG |

| | | |
|---|---|---|
| | 0.983 AA1: 30 AA2: 31 Probability: 0.880 AA1: | |
| 92042 | 30 AA2: 31 Probability: 0.740 AA1: | MKIVKKILIIVSITVFAIFLILLVTPVLFK NSKTHPAALVLRTLALLVFIYQIRLLATDM |
| 92046 | 43 AA2: 44 Probability: 0.902 AA1: | ADTSFFIVSLAGA MKHIPKAAFLIPVLEKLFRLLTHRHKIYLIILFVMTVG |
| 92050 | 57 AA2: 58 Probability: 0.998 AA1: | FALIETLGISVIMPFISMA |
| 92052 | 35 AA2: 36 Probability: 1.000 AA1: | MKTNGKFNQLPLILCSVLLTFFILTCVSVSTSNYA |
| 92069 | 28 AA2: 29 Probability: 0.932 AA1: | MYFSAKRTAAIVLVAVVLCLTVTAVSRA |
| 92103 | 34 AA2: 35 Probability: 1.000 AA1: | MKTKHNLLFNIAQTTVFTILLSVFLTACSNCQDA |
| 92107 | 21 AA2: 22 Probability: 0.999 AA1: | MRKIITKAIALLIAFSTVTLA |
| 92111 | 23 AA2: 24 Probability: 0.679 AA1: | MKNKQKMFGIIAASVAIIGFSFA |
| 92117 | 36 AA2: 37 Probability: 0.991 AA1: | VNQLLLMAQDILSKQVPLGQVALLVLFSLPSVIALS |
| 92129 | 29 AA2: 30 Probability: 0.908 AA1: | MKLLKNLLKSPMFVIGASLFLLTLLIALV MAVLAFLPEQRIKHSVRLILSSSLLFLFSVIAGG |
| 92136 | 44 AA2: 45 Probability: 1.000 AA1: | VNLNAFAPFA |
| 92138 | 20 AA2: 21 Probability: 0.965 AA1: | MKKIALVLVVFLMAATLVFA |
| 92144 | 19 AA2: 20 Probability: 0.998 AA1: | MMVMLFLLGLAVFGLMALA |
| 92164 | 21 AA2: 22 Probability: 0.998 AA1: | MTRYNNRHTFLFALFSFIIFA |
| 92167 | 36 AA2: 37 Probability: 0.999 AA1: | MLNNERSQRRFRLLLFLAVAAIHALLLFFIAFNVKT |
| 92183 | 24 AA2: 25 | MKIKFKLSMLIIAIVAVIVTGVSA |
| 92190 | Probability: | MTTCIMLALVVLGLFSLRALG |

| | | |
|---|---|---|
| | 0.993 AA1: 21 AA2: 22 Probability: | |
| | 0.918 AA1: | |
| 92215 | 21 AA2: 22 | VNTITVCMGSSCFGKGNAANA |
| | Probability: | |
| | 1.000 AA1: | |
| 92221 | 23 AA2: 24 | MKTKAIKFCLTLLIIFFVSPVFA |
| | Probability: | |
| | 0.969 AA1: | |
| 92232 | 19 AA2: 20 | MGLLILLVIIIAFTAAALR |
| | Probability: | |
| | 0.970 AA1: | |
| 92233 | 27 AA2: 28 | KTIAVATLLLLVFIGGFLFLKPSQRRA |
| | Probability: | |
| | 0.895 AA1: | |
| 92240 | 25 AA2: 26 | MIKTVLGFGYLVFAMIFLLPFAITA |
| | Probability: | |
| | 0.998 AA1: | |
| 92241 | 18 AA2: 19 | MIKYLAVFLLFINFTVYA |
| | Probability: | |
| | 0.999 AA1: | |
| 92248 | 35 AA2: 36 | MTMSRVRSLIKFANNKFLVFSCILFLVIALTAITA |
| | Probability: | |
| | 0.959 AA1: | |
| 92261 | 24 AA2: 25 | GVGKSLFAANLGVALAQAGQRVVA |
| | Probability: | |
| | 1.000 AA1: | |
| 92262 | 22 AA2: 23 | MKKAICAVIVLVVIVLVFVLYA |
| | Probability: | |
| | 0.626 AA1: | MGNGKKSKDKFRLVDLLVILICLFGAFYSV |
| 92267 | 43 AA2: 44 | NMFWSDLFQTLNG |
| | Probability: | |
| | 0.868 AA1: | MNKILKEKIMEAFGSVLPITVIVLIVSVILTPIP |
| 92272 | 45 AA2: 46 | SGIFLMFMAGA |
| | Probability: | |
| | 0.989 AA1: | |
| 92286 | 15 AA2: 16 | MNFSIIFAVILLVLA |
| | Probability: | |
| | 1.000 AA1: | |
| 92287 | 25 AA2: 26 | MKKIFVLCAVALFVAGMAFATGSQA |
| | Probability: | |
| | 0.998 AA1: | |
| 92306 | 20 AA2: 21 | MRGYVIITAMVLLMALAVYS |
| | Probability: | |
| | 0.999 AA1: | |
| 92314 | 18 AA2: 19 | MFSRFLILVAFVAGFTFA |
| | Probability: | |
| | 0.990 AA1: | |
| 92317 | 19 AA2: 20 | MKKLIFVLLTVSAIFAVTS |
| 92320 | Probability: | LSIASTSIRRPVLTIVLSALIVLFG |

| | | |
|---|---|---|
| | 0.726 AA1: 25 AA2: 26 Probability: 0.749 AA1: | |
| 92333 | 21 AA2: 22 Probability: 0.959 AA1: | KILVFLGVIALSVVIVRPVQQ |
| 92334 | 29 AA2: 30 Probability: 0.987 AA1: | MNYSNLPVIIAGAFLLGLSAFCSGCYTLK |
| 92336 | 27 AA2: 28 Probability: 0.999 AA1: | RNRMRTFLSVLGIVIGVAAVITMVAIG |
| 92339 | 17 AA2: 18 Probability: 0.942 AA1: | MKRLFLLLILVPAFVFA |
| | | MVNENRSNNNPFPLPFASSAPLRLCERNSR |
| 92351 | 55 AA2: 56 Probability: 0.999 AA1: | RSAPWRWRIFVLCALLAVIGSPVWA |
| 92366 | 30 AA2: 31 Probability: 0.734 AA1: | MKIKIKLSLMVIAIITVVIAVVAVILLRQA |
| 92378 | 20 AA2: 21 Probability: 0.988 AA1: | MRTISGATGRTGTAAGCAAG |
| 92385 | 30 AA2: 31 Probability: 0.925 AA1: | MRTRNTVVMAILASSAVIILIALVIVFFIA |
| 92415 | 28 AA2: 29 Probability: 0.914 AA1: | VIFSRRLTDNIFAWLFRAVSFTGAAALG |
| 92417 | 22 AA2: 23 Probability: 0.993 AA1: | MFTKMKRFLLPIIMFLIFETTA |
| 92436 | 31 AA2: 32 Probability: 0.999 AA1: | MNGNNRRKIVPTVFLVVGVMLMFIDAGAARA |
| 92441 | 25 AA2: 26 Probability: 0.986 AA1: | VLTSRKILFSLLITIVLFTAFTVFA |
| 92452 | 21 AA2: 22 Probability: 1.000 AA1: | MKPLRLSKKNKKIAGVCGAFA |
| 92457 | 24 AA2: 25 Probability: 0.829 AA1: | MIRRKNTAFLAACLIFCLHIPLWA |
| 92464 | 30 AA2: 31 Probability: 0.555 AA1: | VNTSGQASVTPSGSAGATTTYTLTATSALG |
| 92490 | 27 AA2: 28 | MLPISYGAMLVETLVAILALIAVGSLA |
| 92523 | Probability: | MKKIFKVLATQSRWLCAIALVAVIGFSFVT |

| | | |
|---|---|---|
| | 0.988 AA1: | |
| | 30 AA2: 31 | |
| | Probability: | |
| | 1.000 AA1: | MKKKTESTVLKFTQIKSITDVFFNLIMLALLVM |
| 92526 | 47 AA2: 48 | CLVPLLLVISISFS |
| | Probability: | |
| | 0.720 AA1: | GWYMGVVTPRANYYQSTKHLAIILSALGVIFA |
| 92553 | 49 AA2: 50 | VLLGWMLLTISAGKAAS |
| | Probability: | |
| | 1.000 AA1: | |
| 92554 | 24 AA2: 25 | MRQTLLYAAYALSVFCLSALPASA |
| | Probability: | |
| | 0.999 AA1: | |
| 92569 | 18 AA2: 19 | MVKRCIILLFFSIFSACA |
| | Probability: | |
| | 0.988 AA1: | |
| 92573 | 37 AA2: 38 | MNGLINFFIRRPVTVVMILAAIISAAILSLFSMPVNR |
| | Probability: | |
| | 0.708 AA1: | MYRLLSKIKSKNAIIRDGSLEKKYSQFLLHLLRYPVK |
| 92574 | 66 AA2: 67 | VFAAAVLLIIAGFLLVFTRPVVFISPDEA |
| | Probability: | |
| | 1.000 AA1: | |
| 92591 | 35 AA2: 36 | VRILKYVTRKRLIIGGALLVLLLAFIFSFPRPLFN |
| | Probability: | |
| | 0.968 AA1: | |
| 92594 | 29 AA2: 30 | GIQLQIIVTLVFVFLFFFASGNFAPVVRA |
| | Probability: | |
| | 0.846 AA1: | |
| 92600 | 28 AA2: 29 | MIRIDGLAASLKQLLCALCVIIATPLFS |
| | Probability: | |
| | 0.995 AA1: | |
| 92602 | 17 AA2: 18 | MNKIITLLALLNIAVFA |
| | Probability: | |
| | 0.963 AA1: | |
| 92607 | 17 AA2: 18 | VLFILVLALVANTFAQG |
| | Probability: | |
| | 1.000 AA1: | |
| 92620 | 22 AA2: 23 | MKIFFILFAAAVLAGCSTSETA |
| | Probability: | |
| | 0.773 AA1: | |
| 92625 | 22 AA2: 23 | KPVANLAPFGLCKSLANPTVAA |
| | Probability: | |
| | 0.976 AA1: | |
| 92632 | 21 AA2: 22 | DTYRITPLSLWNAASAGLAAA |
| | Probability: | |
| | 0.995 AA1: | |
| 92634 | 23 AA2: 24 | MANKKIWLIMLVIALALGLTVVG |
| | Probability: | |
| | 0.753 AA1: | |
| 92648 | 32 AA2: 33 | MKKSSTFLFCLRLFFYFSVITLILIHPGVTVS |
| 92660 | Probability: | MAQITLIGMLVLGVVALVLIYFLASFFSLWLQ |

| | | |
|---|---|---|
| | 0.947 AA1: 38 AA2: 39 Probability: | ALFSKA |
| | 0.999 AA1: | |
| 92667 | 18 AA2: 19 Probability: | MKKIVVLFFLIAVGGVYA |
| | 0.698 AA1: | |
| 92677 | 28 AA2: 29 Probability: | MNIGIIGLGLIGGSFAKAYKAAGHTVYA |
| | 0.963 AA1: | |
| 92679 | 30 AA2: 31 Probability: | MRLPPIIFIALAFTVMFWSIPSTRVRPSSA |
| | 1.000 AA1: | |
| 92702 | 24 AA2: 25 Probability: | MKRNNSASFLIAVSFSILFTACVA |
| | 0.966 AA1: | |
| 92719 | 17 AA2: 18 Probability: | DTICAAYALFLAQSAFG |
| | 0.594 AA1: | |
| 92743 | 20 AA2: 21 Probability: | MSGLIPILLAVLLLGLVVIG |
| | 1.000 AA1: | |
| 92744 | 23 AA2: 24 Probability: | MKNKLIACLILILMFLSVTSVFA |
| | 0.997 AA1: | |
| 92767 | 23 AA2: 24 Probability: | VKPRPLLPGLSAAALVLALCACA |
| | 0.977 AA1: | |
| 92782 | 23 AA2: 24 Probability: | MKNIKDKLILITGGASGIGQLMA |
| | 0.995 AA1: | |
| 92796 | 24 AA2: 25 Probability: | MKTLTQRLVISAVTLFAGLGLSTA |
| | 0.918 AA1: | |
| 92802 | 35 AA2: 36 Probability: | VKETQDNEKKTSIKVRYIIFSISLFLIIFIGGSAA |
| | 0.959 AA1: | |
| 92806 | 25 AA2: 26 Probability: | MRGLKKKHNRRFTLVVAGCMAQRLG |
| | 0.960 AA1: | |
| 92812 | 19 AA2: 20 Probability: | MGKLKLCTLVLALAFFAIG |
| | 1.000 AA1: | |
| 92820 | 26 AA2: 27 Probability: | LSLSKTGASRLLLGLIILLIVLGAAA |
| | 0.991 AA1: | |
| 92829 | 37 AA2: 38 Probability: | MNIKQSVNNFKTRVLSSYIMIPAVILALLLIAAISNN |
| | 1.000 AA1: | MNIKQAVKTIKVRVLNSYILVLVVVLAFALMVALSSYF |
| 92830 | 53 AA2: 54 | MTDIVKRQLASNAHA |
| 92892 | Probability: | VNRKNAPDYRPLMLILLVFLGVSMLPA |

| | | |
|---|---|---|
| | 0.997 AA1: 27 AA2: 28 Probability: 0.999 AA1: | |
| 92900 | 21 AA2: 22 Probability: 1.000 AA1: | MSSLAAGLVLGSCAGWGAASA MTKKNKALAAVIACVLFWGFSFISIKIAVAVFPPM |
| 92902 | 45 AA2: 46 Probability: 0.963 AA1: | TLGALRFALA |
| 92913 | 23 AA2: 24 Probability: 0.810 AA1: | MWFLIFSTGVLLIFLLNVSNGSA |
| 92936 | 35 AA2: 36 Probability: 0.997 AA1: | MVLSAPPEFKLRKREGAIAALALFMLVFFGGALFV |
| 92955 | 25 AA2: 26 Probability: 0.968 AA1: | MKRLDWLFMISLFALIGIGLALVYS |
| 92978 | 23 AA2: 24 Probability: 1.000 AA1: | MTKKKLLIGLLAVMLILGGCPES |
| 92989 | 31 AA2: 32 Probability: 0.999 AA1: | VFLVNKVSRRIRTLAMVLLFASAAASQEALA |
| 92991 | 21 AA2: 22 Probability: 0.991 AA1: | MNFRPFALILVFLFLSAVIFA |
| 93005 | 33 AA2: 34 Probability: 0.729 AA1: | MIRKIAVCLTMFALGSFFVGCNDKGGVSGPVSG |
| 93009 | 29 AA2: 30 Probability: 1.000 AA1: | MKNEKKSKDKFRLVDFLVILICLFGALYS |
| 93027 | 28 AA2: 29 Probability: 0.996 AA1: | VVMKIDIKFLVVGFSFLMFLFLAATVFA LKRRLDLFISAIAGVYASATANLFTSAKACRT |
| 93062 | 45 AA2: 46 Probability: 0.634 AA1: | AVALIFAASSLFA |
| 93069 | 25 AA2: 26 Probability: 0.939 AA1: | KIASIKNNLPVYVFSGSSDPVGAMG |
| 93073 | 20 AA2: 21 Probability: 0.949 AA1: | MNKFMFVSGGVCSSLGKGVA |
| 93077 | 25 AA2: 26 Probability: 0.637 AA1: | MAVKNLGIAVFAVMFVFALAGCNEA |
| 93078 | 32 AA2: 33 | MTILTLLLRMTGGLCMFLFGMKMMSDGLQKSA |
| 93106 | Probability: | FWVIAITAMIVSVLTGCVSYSVQSVVISAANN |

| | | |
|---|---|---|
| | 0.935 AA1: 32 AA2: 33 Probability: | |
| | 0.995 AA1: | |
| 93111 | 37 AA2: 38 | MGSRSNSATSTWLVRLLAAFTVILAVFIGIGLGLSLA |
| | Probability: 0.891 AA1: | |
| 93114 | 41 AA2: 42 | LNVFSFGSAQYIKNKKNSVIITALIAVITFVFFTGNSILSK |
| | Probability: 1.000 AA1: | |
| 93116 | 22 AA2: 23 | MKLKPPIMAFLLFFAGITAVYS |
| | Probability: 0.849 AA1: | MSKFLTALTRLAGSAGSAKLMKLAGLTVLT |
| 93122 | 41 AA2: 42 | ALLLSCSDDT |
| | Probability: 1.000 AA1: | |
| 93149 | 34 AA2: 35 | LKFKYLIIAFSILIVLFLLTTTLFPTLIPGSVYA |
| | Probability: 0.980 AA1: | |
| 93152 | 21 AA2: 22 | MKRIGFIVLALLLCGAMVLTG |
| | Probability: 1.000 AA1: | |
| 93168 | 29 AA2: 30 | LGKKSLKFFEKSQGVFVFLLLLIGARVFA |
| | Probability: 0.999 AA1: | |
| 93171 | 20 AA2: 21 | MKIKFKLSIVVIAIMAVVVA |
| | Probability: 0.999 AA1: | |
| 93181 | 31 AA2: 32 | MKNILKKNYIFVAAAILIVSAYFITGVLVFA |
| | Probability: 0.997 AA1: | |
| 93182 | 46 AA2: 47 | VNYLKKRKITDSFWYALMTAASIAVGAALVGIIAYILIQALGMSAA |
| | Probability: 0.770 AA1: | |
| 93188 | 23 AA2: 24 | SYPFSGNGYGGFTLFVSSSSVLS |
| | Probability: 0.876 AA1: | |
| 93189 | 35 AA2: 36 | MDRYGVAPKKIVYAIILCIIIFSGCASVKSAPQQA |
| | Probability: 1.000 AA1: | |
| 93195 | 22 AA2: 23 | MKIRLKMFLVILPLIIVPLVLA |
| | Probability: 0.720 AA1: | |
| 93226 | 24 AA2: 25 | MHSAQIYFLVTALSYPFLALYTAA |
| | Probability: 0.996 AA1: | |
| 93262 | 24 AA2: 25 | MDKLLLMILAVSFSACSSTHVRSG |
| | Probability: 0.978 AA1: | |
| 93278 | 21 AA2: 22 | VKKHPEFFVLLILMLGFPLFS |
| 93304 | Probability: | MKIRFKIILVVLPVVILALCIA |

| | | |
|---|---|---|
| | 0.999 AA1: 22 AA2: 23 Probability: 1.000 AA1: | |
| 93327 | 33 AA2: 34 | MKFTMRTTLNTIRFRRLIVSCILFALCVSALWA |
| | Probability: 0.950 AA1: | |
| 93334 | 34 AA2: 35 | SFSGREISLLGGVMVVFSFIAAIIATEMIAKAYA |
| | Probability: 0.939 AA1: | |
| 93337 | 29 AA2: 30 | MRKLYTYPWLIVAITALITVFFAFQLPHA |
| | Probability: 1.000 AA1: | |
| 93338 | 29 AA2: 30 | MINSRKIINIILSVLFISMTVSSCGKSTA |
| | Probability: 0.996 AA1: | |
| 93361 | 33 AA2: 34 | MNLFKNASISKKLIMLLVVFILGYSIFGLVAFS |
| | Probability: 0.997 AA1: | |
| 93367 | 26 AA2: 27 | MNKRIIFFISLLIISGLIFCACISNG |
| | Probability: 0.999 AA1: | |
| 93371 | 27 AA2: 28 | LNFILRRIVIALITVFLVSIFCFLAFS |
| | Probability: 0.541 AA1: | |
| 93402 | 30 AA2: 31 | LSIASTSVRRPVLTLVLSAFIVLFGLIGLK |
| | Probability: 0.667 AA1: | |
| 93420 | 21 AA2: 22 | MSPLRWLLLPLAPLYAIAVRA |
| | Probability: 0.995 AA1: | |
| 93422 | 15 AA2: 16 | MFTVLITLAAAAATA |
| | Probability: 0.810 AA1: | |
| 93459 | 31 AA2: 32 | MDDLRRKNSLDFTLLLTAVALWVIGIFLVYS |
| | Probability: 0.748 AA1: | MQIKKILFVNKRLKTASDGVYALALLLPYGLLF |
| 93466 | 44 AA2: 45 | SVFIAIPIAVA |
| | Probability: 0.951 AA1: | |
| 93472 | 23 AA2: 24 | VAPYKARVIAVAVFLVIFFLAVA |
| | Probability: 1.000 AA1: | |
| 93490 | 20 AA2: 21 | MKKTGFIVVAFLLCGAMVFA |
| | Probability: 0.907 AA1: | |
| 93495 | 37 AA2: 38 | MSNEQLTINNEKRAMSVNLRLLLVSLLLFGLCWNVSA |
| | Probability: 0.813 AA1: | |
| 93505 | 19 AA2: 20 | MPVFFVILLVLFNVFPVYS |
| 93508 | Probability: | KMHFLPGHLHWITALTAAMLAGALWG |

| | | |
|---|---|---|
| | 0.927 AA1: 26 AA2: 27 Probability: 1.000 AA1: | |
| 93517 | 21 AA2: 22 Probability: 1.000 AA1: | MKFLRRAIIVILIFIPFYAFG |
| 93521 | 20 AA2: 21 Probability: 0.976 AA1: | MRKLILTTAAVAAAIAGAFA |
| 93522 | 31 AA2: 32 Probability: 0.966 AA1: | MRKFLTPQAVIITVVVVVLALFFLLAGKQNA |
| 93524 | 33 AA2: 34 Probability: 0.979 AA1: | LRPEVMQKRFIYSLYIAVILWSAVSCASGKTAA |
| 93561 | 28 AA2: 29 Probability: 0.995 AA1: | MKQKHGLTDSNLLFLIAVCIFVILYALA LANKKKKKIVKSLIHLFTVLVALVIIFVPILI |
| 93568 | 44 AA2: 45 Probability: 0.999 AA1: | GVVMYFNAPLHQ |
| 93580 | 26 AA2: 27 Probability: 0.633 AA1: | VKSLKSYLTFAAAVILLAALWLLVVS |
| 93583 | 36 AA2: 37 Probability: 0.739 AA1: | MTFNEKMHSPLWIFIIYMIVSCSVIMIFRFVFPGSA |
| 93612 | 20 AA2: 21 Probability: 0.781 AA1: | MHDILVIGAGIAGLTVALKA |
| 93620 | 26 AA2: 27 Probability: 0.909 AA1: | MKTFLNMLKILVLGIQIIACSSGSYE |
| 93645 | 30 AA2: 31 Probability: 0.541 AA1: | MSMVKTVVSRPTTFFIIFALLIGLGFFALV |
| 93656 | 29 AA2: 30 Probability: 0.997 AA1: | MEMNNKKIIVKLPLILTAFIGLFMLACEK |
| 93666 | 39 AA2: 40 Probability: 0.714 AA1: | MPTFYELFKKIKSVRSLFLQLLFVALAFILMILSSSLYV |
| 93668 | 22 AA2: 23 Probability: 0.983 AA1: | ASTGSVTFTGGTVSMNSQGAYS |
| 93695 | 18 AA2: 19 Probability: 0.966 AA1: | MKSSKIPLIIIASLIALS |
| 93704 | 30 AA2: 31 | MSMVKTVVSRPTTFLIIFALLIGLGFFAMV |
| 93705 | Probability: | MGMVNTVVSRPTTFFIIFALLIGLGFFALV |

| | | |
|---|---|---|
| | 0.537 AA1: 30 AA2: 31 Probability: | |
| | 0.979 AA1: | MVKNIQLTPVLCAATGAAAGFYLFGRLSFAPL |
| 93719 | 41 AA2: 42 Probability: | FVIAFVLLA |
| | 0.625 AA1: | |
| 93720 | 24 AA2: 25 Probability: | MMNTKGSSGPAGFLSGVFNSPVLQ |
| | 0.999 AA1: | |
| 93724 | 22 AA2: 23 Probability: | MINKKKCFVFLILFCLYHGLQA |
| | 0.949 AA1: | MVFFLVLRNIAGNKKNCAIIALLIAVITF |
| 93725 | 39 AA2: 40 Probability: | LFFIGNSVIG |
| | 1.000 AA1: | |
| 93731 | 21 AA2: 22 Probability: | MTMKNRILILLLTGLCAAAFA |
| | 0.958 AA1: | |
| 93733 | 19 AA2: 20 Probability: | VHAPAGSIVMYATSAGSVA |
| | 0.795 AA1: | |
| 93735 | 30 AA2: 31 Probability: | VRKLFLFFILFSLFPIPYSPLPTPHSLLFA |
| | 0.806 AA1: | |
| 93741 | 16 AA2: 17 Probability: | MRTFTGLPVSTGVAYA |
| | 1.000 AA1: | |
| 93762 | 21 AA2: 22 Probability: | MKRKLFIFSALFVFASFSAFA |
| | 0.797 AA1: | |
| 93784 | 22 AA2: 23 Probability: | MKNRQIFPVVILFFLVSLRLFG |
| | 0.562 AA1: | |
| 93790 | 19 AA2: 20 Probability: | MKLPLSHGSITVQVANAQA |
| | 0.997 AA1: | |
| 93791 | 26 AA2: 27 Probability: | MIVQYHKFALTAVFGLALAFTFSCTS |
| | 0.925 AA1: | |
| 93796 | 27 AA2: 28 Probability: | LSTVWRLFLVTLAGVLMSLNIKTFVHA |
| | 0.999 AA1: | |
| 93802 | 26 AA2: 27 Probability: | LNKLNRLLFLFSLFFIAALSSCSKLG |
| | 0.997 AA1: | |
| 93825 | 21 AA2: 22 Probability: | GRALAFAAIALGTIAALPAIA |
| | 0.548 AA1: | |
| 93826 | 24 AA2: 25 | MPEKFILGLGIPFPGCLAGALVCA |
| 93832 | Probability: | LQAIFAGQLAAAPLDSGFALAPKASA |

| | | |
|---|---|---|
| | 0.756 AA1: | |
| | 26 AA2: 27 | |
| | Probability: | |
| | 0.513 AA1: | |
| 93840 | 21 AA2: 22 | MDTITVCMGSSCFGKGNSANA |
| | Probability: | |
| | 0.618 AA1: | MKTGREKQQNRLGSDFYVFVILILVSFSAIF |
| 93862 | 39 AA2: 40 | FSSRNVLG |
| | Probability: | |
| | 1.000 AA1: | |
| 93884 | 28 AA2: 29 | MKIAFNFRVAGFLFTAVLFLAAAFTAFA |
| | Probability: | |
| | 0.585 AA1: | |
| 93890 | 21 AA2: 22 | MKYSNPANLAVVACPGGEIFA |
| | Probability: | |
| | 0.999 AA1: | |
| 93892 | 23 AA2: 24 | MRKNLIKSLFFGIPALTAFIAFA |
| | Probability: | |
| | 0.996 AA1: | |
| 93894 | 23 AA2: 24 | MNLKKRYIFIFSCLAVFAVYLLA |
| | Probability: | |
| | 1.000 AA1: | |
| 93899 | 21 AA2: 22 | MKQVKNLFGVAIALAIVVSMA |
| | Probability: | |
| | 0.996 AA1: | |
| 93901 | 20 AA2: 21 | MRGKLLLLCVLLLVVENLFA |
| | Probability: | |
| | 1.000 AA1: | |
| 93907 | 21 AA2: 22 | MKKKTVLALLMATMLTGGVFA |
| | Probability: | |
| | 0.998 AA1: | |
| 93924 | 33 AA2: 34 | VKLPSIFRFRNLCIAFGAAFIVVVLCAVLAIYA |
| | Probability: | |
| | 0.969 AA1: | |
| 93927 | 26 AA2: 27 | MNRKLFVILIVAVIAIFGLGSCLTNR |
| | Probability: | |
| | 0.976 AA1: | |
| 93942 | 22 AA2: 23 | MKKFQTVIIFVIAAFFSACTTV |
| | Probability: | |
| | 0.658 AA1: | |
| 93947 | 30 AA2: 31 | MRIFPVLLIFLASSIFSLSFFDDRPPEELA |
| | Probability: | |
| | 1.000 AA1: | |
| 93954 | 22 AA2: 23 | MIRSIFFSAALLLLVSTQNLFA |
| | Probability: | |
| | 0.991 AA1: | |
| 93957 | 20 AA2: 21 | MTVLPLTLVMIILAAIAAQA |
| | Probability: | |
| | 0.929 AA1: | |
| 93965 | 29 AA2: 30 | MKKINFPIICRILAALAVVMYMCLFTACK |
| 93987 | Probability: | VIHPNRVFQKKYLSLAVTLLLCLLSVCQKSVV |

| | | |
|---|---|---|
| | 0.992 AA1: | |
| | 32 AA2: 33 | |
| | Probability: | |
| | 0.996 AA1: | |
| 93988 | 23 AA2: 24 | MKKSNLWTGILAITLVFTMAITA |
| | Probability: | |
| | 1.000 AA1: | |
| 93992 | 20 AA2: 21 | MKKTAFLALIFYAIAANGYA |
| | Probability: | |
| | 0.997 AA1: | |
| 93993 | 34 AA2: 35 | MKIKFKLSILIIAIVVVIVTGVAVLLLRQASSIA |
| | Probability: | |
| | 1.000 AA1: | |
| 93994 | 19 AA2: 20 | MKKIFFVICLLFSFSFCFA |
| | Probability: | |
| | 0.901 AA1: | |
| 93996 | 23 AA2: 24 | MKIGLKLTVLMTVLGVFTVASIS |
| | Probability: | |
| | 0.999 AA1: | |
| 93997 | 30 AA2: 31 | MKIRIKLSLMVIAIVSVVAGGIAIILLLQA |
| | Probability: | |
| | 1.000 AA1: | |
| 93999 | 25 AA2: 26 | MKKNFCLLIAFCLLLFSQCSSKASA |
| | Probability: | |
| | 0.942 AA1: | |
| 94027 | 21 AA2: 22 | MKRKVVVTGMGAISSIGKSVA |
| | Probability: | VASALAAALITLISLIATAFTLVALIALV |
| | 1.000 AA1: | SLITFAALVA |
| 94049 | 58 AA2: 59 | AAFTLAAALIALVSLITFA |
| | Probability: | |
| | 0.835 AA1: | |
| 94064 | 22 AA2: 23 | VASKETTTIVALATPPGTSAIA |
| | Probability: | |
| | 0.503 AA1: | |
| 94071 | 18 AA2: 19 | MVSAILLMIIYLSFISLG |
| | Probability: | |
| | 0.513 AA1: | |
| 94073 | 21 AA2: 22 | MDTITVCMGSSCFGKGNSANA |
| | Probability: | |
| | 0.698 AA1: | |
| 94077 | 19 AA2: 20 | LKNLWSALCKLSLPTGASG |
| | Probability: | |
| | 0.915 AA1: | |
| 94120 | 22 AA2: 23 | MTRKILRILTIIIFIVCIPGIA |
| | Probability: | |
| | 1.000 AA1: | |
| 94132 | 19 AA2: 20 | MKKRILIFILLIVSSFLYA |
| | Probability: | |
| | 0.743 AA1: | KRWVRNLSRNISAIRDYAGISRLAGLAA |
| 94148 | 36 AA2: 37 | DGAPVFLA |
| 94151 | Probability: | AGVMGTDCKVVKVSGANVKMFGDIKIV |

| | | |
|---|---|---|
| | 0.741 AA1: | FAAAFIAA |
| | 45 AA2: 46 | AFASQGFAQA |
| | Probability: | |
| | 0.603 AA1: | |
| 94153 | 21 AA2: 22 | MKYFNPANLAVIACPGGEIFA |
| | Probability: | MTPKEKTLINKLFTVSFPALFKNRSILAW |
| | 0.980 AA1: | VGAFLFCL |
| 94184 | 48 AA2: 49 | VVILLSGNSSS |
| | Probability: | |
| | 0.775 AA1: | |
| 94195 | 32 AA2: 33 | MIFGFLAAAAGVYSVLIFIRIIISWFGGSVYG |
| | Probability: | |
| | 0.794 AA1: | |
| 94217 | 29 AA2: 30 | SAFEEKLKKGLSAPLVMTLFFISAFMVLM |
| | Probability: | |
| | 0.723 AA1: | |
| 94224 | 22 AA2: 23 | YTEASGLSAGTLTIAKATPVLA |
| | Probability: | |
| | 0.667 AA1: | |
| 94234 | 19 AA2: 20 | RSRAERHRVRAALSVFSWA |
| | Probability: | MLEKRKIKTNQVSGNAAMPRKKRRLKYIAVLPS |
| | 0.795 AA1: | FITLMNGVCGFFAIIIASRSPDIRWNFSFIPGAGMS |
| 94236 | 74 AA2: 75 | PFTLA |
| | Probability: | |
| | 0.994 AA1: | |
| 94237 | 19 AA2: 20 | LYGLCLILLFTSVCFNIQA |
| | Probability: | |
| | 1.000 AA1: | |
| 94239 | 26 AA2: 27 | MNRKTITVKTAAFAFLLAVCASAAYA |
| | Probability: | |
| | 0.936 AA1: | |
| 94245 | 26 AA2: 27 | AYTGCKKKFDTTGSVATPIFASATFA |
| | Probability: | |
| | 0.992 AA1: | VPDKKIIFLNKPQINISCPSALRMWLVCACAFL |
| 94258 | 41 AA2: 42 | VIQSAFS |
| | Probability: | |
| | 0.757 AA1: | RASLARRCLALALAATLLIICGFIPFWRRVSRLLFDA |
| 94262 | 59 AA2: 60 | EDLWAVLLLAASLFLSVTCAFA |
| | Probability: | |
| | 1.000 AA1: | |
| 94267 | 22 AA2: 23 | MKKLFVLIALVYLFLPLSFSSA |
| | Probability: | |
| | 0.999 AA1: | |
| 94270 | 21 AA2: 22 | MKNKRFLMAVLAMLTVFALIG |
| | Probability: | |
| | 0.999 AA1: | |
| 94283 | 20 AA2: 21 | MKNIAAMIVLAAMCSVGGCA |
| | Probability: | |
| | 0.999 AA1: | |
| 94290 | 17 AA2: 18 | MKKTILYAAAITTAVTA |
| 94309 | Probability: | MKQVCIFLLFLLLALSSCTKN |

| | | |
|---|---|---|
| | 0.995 AA1: 21 AA2: 22 Probability: 1.000 AA1: | |
| 94322 | 25 AA2: 26 Probability: 0.887 AA1: | VLTSRKILLSLLITIILFTAFTVFA |
| 94323 | 29 AA2: 30 Probability: 0.999 AA1: | HLKRQRQVQRKKRERSGIPVCALVGYTNA |
| 94344 | 27 AA2: 28 Probability: 0.999 AA1: | MIKIKKRKIIDGLIYGIMAAAMTAVLA |
| 94346 | 34 AA2: 35 Probability: 1.000 AA1: | MYNRTQSAYRKALRRKRLIFTAAVLVFVGF AVFA |
| 94354 | 32 AA2: 33 Probability: 0.547 AA1: | MYFRRKYILLNIILLLSPGFLLFQSCATGSFA |
| 94361 | 18 AA2: 19 Probability: 0.881 AA1: | MKKRISILGATGSIGKSA |
| 94384 | 19 AA2: 20 Probability: 0.895 AA1: | MKLNTLLFIGTAAILVAGG |
| 94391 | 36 AA2: 37 Probability: 0.999 AA1: | IYWISAAVLMCFALLPRSPLSLRVVLVLLGALMAFA |
| 94396 | 23 AA2: 24 Probability: 0.978 AA1: | MKNIIKQICVIAFVAIIGFSFFA MLSIAYVGFSVIAGIFLGRFGIKPSFLFGFSALCTGLF |
| 94398 | 71 AA2: 72 Probability: 1.000 AA1: | SMFFMPGFFLTAAALFVVFAGFGFFEVGLNALA |
| 94413 | 21 AA2: 22 Probability: 0.601 AA1: | MKIRIFLILIFAFFSLLPLQA |
| 94417 | 16 AA2: 17 Probability: 1.000 AA1: | MKTFTGIPVSTGVAYA VTKKRQFVRILKYVTKKRLIIVGALLVLLLAFI |
| 94419 | 42 AA2: 43 Probability: 1.000 AA1: | FSLPRPLFN |
| 94420 | 35 AA2: 36 Probability: 1.000 AA1: | VRILKYVTKKRLIIVGALLVLLLAFIFSLPRPLFN |
| 94430 | 20 AA2: 21 Probability: 0.999 AA1: | MKRNLFFVFFSLVFSTIVSA |
| 94431 | 27 AA2: 28 | MIRRIIFIFLTGLLLSTCAGQGQPAIS |
| 94433 | Probability: | VKKVAVVITVLCVLMAFNACG |

| | | |
|---|---|---|
| | 0.994 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 0.982 AA1: | |
| 94438 | 22 AA2: 23 | MKNQKYTFLALILLTGFAMIAA |
| | Probability: | |
| | 0.962 AA1: | MNLPKLSVTRRVTVTMLILIMVLFGALSF |
| 94441 | 33 AA2: 34 | VRLG |
| | Probability: | |
| | 0.992 AA1: | |
| 94457 | 23 AA2: 24 | MRKFVLHITAVLIYFCVVFSVHA |
| | Probability: | |
| | 0.676 AA1: | |
| 94461 | 25 AA2: 26 | MLKLKSKFIFVTGGVVSGLGKGITA |
| | Probability: | |
| | 0.974 AA1: | |
| 94468 | 24 AA2: 25 | VSTLKKTALVLIISGLLYAGMVYL |
| | Probability: | |
| | 0.969 AA1: | |
| 94472 | 30 AA2: 31 | MSRKQKIIIISAGVALTALIVLGIIFFVKG |
| | Probability: | |
| | 1.000 AA1: | |
| 94477 | 23 AA2: 24 | MKKYTLFFVCCLISLIAAATVSA |
| | Probability: | |
| | 0.605 AA1: | |
| 94478 | 29 AA2: 30 | TPPNTLFMANMRLILISIICLLLLPSCSK |
| | Probability: | |
| | 0.920 AA1: | |
| 94481 | 30 AA2: 31 | ARLLEARQLIAVRRYSAALASLAIFQPVAV |
| | Probability: | |
| | 0.623 AA1: | |
| 94489 | 24 AA2: 25 | MKKVKLAFMIIFDLFIILSCTTSV |
| | Probability: | |
| | 1.000 AA1: | |
| 94502 | 27 AA2: 28 | MGLINRYWRALSLAVAATALYAFNAAA |
| | Probability: | |
| | 0.998 AA1: | |
| 94505 | 22 AA2: 23 | MKNFGKLLGIISLVAIIVFAMA |
| | Probability: | |
| | 1.000 AA1: | |
| 94525 | 23 AA2: 24 | MKKSLKKIAFSFVVLLLAACQSA |
| | Probability: | |
| | 0.621 AA1: | |
| 94541 | 28 AA2: 29 | VVSCITAIFNGYVVAALVILVLYSAVGA |
| | Probability: | |
| | 0.960 AA1: | |
| 94542 | 19 AA2: 20 | MIIVLITAVFALAIAFVLG |
| | Probability: | |
| | 1.000 AA1: | |
| 94580 | 20 AA2: 21 | MRKLLLTTAAVAAAIAGAFA |
| 94590 | Probability: | MMAGLGALLSDLSGALAVFIEGFMVIGSF |

| | | |
|---|---|---|
| | 0.940 AA1: 52 AA2: 53 Probability: | FSWIFADWTGSVFFGSLLSAALA |
| | 1.000 AA1: | |
| 94600 | 23 AA2: 24 Probability: | MKKRMLLTGMFALMLVFAMFLTG |
| | 0.998 AA1: | |
| 94606 | 26 AA2: 27 Probability: | LRKKFSYIVFAAALMMPAIVSCSFFG |
| | 0.995 AA1: | |
| 94615 | 19 AA2: 20 Probability: | MSSILIPLVVFFISISALA |
| | 0.929 AA1: | |
| 94621 | 29 AA2: 30 Probability: | MPFTANNLWLFVLAILFMTINGPTQMSYA |
| | 0.989 AA1: | |
| 94624 | 26 AA2: 27 Probability: | MEFSSFDTMWVLIAAALVFFMQAGFA |
| | 0.999 AA1: | |
| 94629 | 27 AA2: 28 Probability: | MNKKLFSINKIICLFFIALCFMFVTCS |
| | 0.992 AA1: | |
| 94634 | 27 AA2: 28 Probability: | VNYLKKRKITDGFWYALMTAASVAVGA |
| | 0.613 AA1: | |
| 94645 | 44 AA2: 45 Probability: | MIADNILLIPLVVIVVLGMLFFSFRRFTFVILPFLTVIIAVVWA |
| | 0.543 AA1: | |
| 94656 | 32 AA2: 33 Probability: | MYIKIPGILKKRPLYVQILFTVFAFLLMVVLS |
| | 0.978 AA1: | |
| 94658 | 20 AA2: 21 Probability: | MKHILKVLIFPLILTGLVYG |
| | 0.874 AA1: | |
| 94659 | 21 AA2: 22 Probability: | LPIKFFFATAFLIILGFSSQL |
| | 0.924 AA1: | |
| 94666 | 20 AA2: 21 Probability: | MKSVILVILLMFFWTNCVYG |
| | 0.608 AA1: | |
| 94677 | 36 AA2: 37 Probability: | MRVKEDGMKIKIKLSIVVIAIMAVVVTGIATLLLWQ |
| | 0.925 AA1: | |
| 94678 | 36 AA2: 37 Probability: | MDETSIYYISITAALVILFALSMFFAFCETSFSSLS |
| | 0.741 AA1: | |
| 94691 | 17 AA2: 18 Probability: | MNGLLLLLISVITLLFA |
| | 0.996 AA1: | |
| 94699 | 19 AA2: 20 | MPVFAAIGGIALIMLQALG |
| 94701 | Probability: | MKTKALLAVITAAFLLSAFTIDSAYA |

| | | |
|---|---|---|
| | 1.000 AA1: 26 AA2: 27 Probability: 0.983 AA1: | |
| 94707 | 22 AA2: 23 | MKKVVFTAALIISVMLAVSCKS |
| | Probability: 0.998 AA1: | |
| 94717 | 30 AA2: 31 | MGGNMREQIIIKKAAFILIALFLLSTGLYA |
| | Probability: 0.853 AA1: | |
| 94733 | 26 AA2: 27 | KGGIPVPLSWLAGLLLWAYDSLPSVA |
| | Probability: 0.640 AA1: | |
| 94737 | 22 AA2: 23 | KLLLAATIISMILGCLSKRVWA |
| | Probability: 0.641 AA1: | |
| 94742 | 37 AA2: 38 | MVFSWMPQLVKGARVTIALTVLAVSAGLVLSLFLALG |
| | Probability: 0.995 AA1: | |
| 94747 | 31 AA2: 32 | LRGICKKTAVLLMLMGLVCTPGFFLFGQNNA |
| | Probability: 0.964 AA1: | |
| 94750 | 20 AA2: 21 | MKTTLPILAIGFCLSILACA |
| | Probability: 1.000 AA1: | |
| 94763 | 37 AA2: 38 | MIKNHVKSLKKIVNKTFLVFSGILFLIIALSAIVAYA |
| | Probability: 0.986 AA1: | |
| 94768 | 30 AA2: 31 | VTHQKIKLLKHATIIPLFGSSISFSGGVYT |
| | Probability: 0.508 AA1: | |
| 94785 | 24 AA2: 25 | MIKTKFLSTAIGSMPFEDPGYAVG |
| | Probability: 0.981 AA1: | |
| 94790 | 28 AA2: 29 | MKRIIRFSNLFIPAAIFSILLTVSGVVS |
| | Probability: 1.000 AA1: | |
| 94796 | 31 AA2: 32 | MKLIKKLLGIAAITAVIGFITLPLTGCPAEA |
| | Probability: 1.000 AA1: | |
| 94797 | 31 AA2: 32 | MKLIKKLLGIAAITAVIGFITLPLTGCPAEA |
| | Probability: 0.754 AA1: | |
| 94808 | 25 AA2: 26 | VRYLRGAAAGIAVSLVPIIVTLIVA |
| | Probability: 0.961 AA1: | |
| 94810 | 18 AA2: 19 | MNKIKAIAVSLVFFTLVA |
| | Probability: 0.572 AA1: | |
| 94816 | 30 AA2: 31 | MRINLKKIIIIAVFFFSALPLFCRDITVLV |
| 94830 | Probability: | MTIKVKLSLMMIAIMVVVA |

| | | |
|---|---|---|
| | 0.993 AA1: 19 AA2: 20 Probability: 0.871 AA1: | |
| 94839 | 25 AA2: 26 Probability: 0.989 AA1: | MKKIVFIALVLLICCGTFLFIGCRK |
| 94841 | 26 AA2: 27 Probability: 1.000 AA1: | IFPRAIKIIASFIIPVAMIGFFPASA |
| 94858 | 19 AA2: 20 Probability: 0.524 AA1: | MLHKAFTILLCTVSLAAHA |
| 94891 | 24 AA2: 25 Probability: 0.971 AA1: | LFFTTIYIAISFIVLFLGTSQIVG |
| 94893 | 33 AA2: 34 Probability: 1.000 AA1: | MKRNQLAALFFISTAGIAFELYVMRIFSVGSWA |
| 94899 | 39 AA2: 40 Probability: 0.957 AA1: | MEKKKLLLVAISVGVFLALTIGAALLVFAPKDTSAASMA |
| 94913 | 34 AA2: 35 Probability: 0.635 AA1: | MSENKSSFFRIFANGIIRENPLLMLMIGLCSSLA |
| 94931 | 21 AA2: 22 Probability: 1.000 AA1: | MTIRQKIGQCFGIGFSGTSVS |
| 94933 | 35 AA2: 36 Probability: 0.967 AA1: | MSLKIKLTIIILVMILVSVAVLSVFTLTRSASLQS |
| 94952 | 30 AA2: 31 Probability: 0.902 AA1: | MTIVLTTALLSLAIAFILGLALGFFKEFFA |
| 94960 | 25 AA2: 26 Probability: 0.760 AA1: | MKRSWKVLIVLVIISIAGGAFYFYK |
| 94971 | 22 AA2: 23 Probability: 0.586 AA1: | MKKIFLILIALSVLTLSCCDKG |
| 94985 | 29 AA2: 30 Probability: 0.958 AA1: | MESGKKPGNKLRLADVLVVLFCLSGAAYS |
| 94986 | 35 AA2: 36 Probability: 1.000 AA1: | AVRFSIGTKLITIISIIVLISLGSITALVSWLVHG |
| 95006 | 38 AA2: 39 Probability: 0.999 AA1: | MAGVFFKSDGKFRLFSVFFSLILLVSMFSGCAGLAASA |
| 95016 | 34 AA2: 35 | MKIKIKLSLMIIAIVAVIVTGVAVLLLRQASSIA |
| 95037 | Probability: | LLNKKILVLCSSVFFLSGFLYA |

| | | |
|---|---|---|
| | 1.000 AA1: 22 AA2: 23 Probability: 0.944 AA1: | |
| 95042 | 31 AA2: 32 | MDNGMKQNQRKKFFIFTAVFYMLVGAACGHA |
| | Probability: 0.840 AA1: | |
| 95053 | 41 AA2: 42 | MPEQKIAYRNKPQINLSYPSTGRMWLVSVCAFLCALQSALT |
| | Probability: 1.000 AA1: | |
| 95080 | 31 AA2: 32 | MSKKINRSIVNFHFLALAAIIIFTAAGNLAA |
| | Probability: 0.993 AA1: | |
| 95114 | 37 AA2: 38 | MKLFHSLRFRFAFIFSVFIIILSVSLSLLGIRQTTKA |
| | Probability: 1.000 AA1: | |
| 95140 | 20 AA2: 21 | MRKFFGIIVCFSLIILHAQA |
| | Probability: 0.982 AA1: | |
| 95167 | 34 AA2: 35 | MNNLKVSMKLLVGFGVMVALILTVSLVSISNLLG |
| | Probability: 0.996 AA1: | |
| 95169 | 25 AA2: 26 | KKRPWYKAAIAASGGVAATTPYVDA |
| | Probability: 0.609 AA1: | |
| 95187 | 45 AA2: 46 | MFRLFRDGYHAFALFAGAATCFVTLIMLRDSLTAWRWLAAGLVMA |
| | Probability: 0.999 AA1: | |
| 95189 | 32 AA2: 33 | LKIDKINTLAFAILCTFGVHTSFAQCTASLFA |
| | Probability: 0.599 AA1: | |
| 95192 | 36 AA2: 37 | YSRLLGALTAIIVRITQKESTAFILILAAIAVSVLV |
| | Probability: 0.855 AA1: | |
| 95199 | 16 AA2: 17 | MPIYLLFGILVSSVIA |
| | Probability: 0.999 AA1: | |
| 95212 | 20 AA2: 21 | MSLAHIKIGFALLLAGTAYA |
| | Probability: 0.977 AA1: | MNLGTLIGVLGAMTMVIGSILLSGNSIM |
| 95222 | 61 AA2: 62 | AFVDIPSLALVLVGSYFALMIAFSLSDALSVFA |
| | Probability: 0.988 AA1: | |
| 95225 | 42 AA2: 43 | MKLLKNLLKSPMFVIGASIFLLTLLIALVGPFFYKITTARVG |
| | Probability: 0.972 AA1: | |
| 95230 | 41 AA2: 42 | MIKQKFLKPPWLEMIIMLAAVALSAVGLLIISYLSRPNISA |
| | Probability: 0.996 AA1: | |
| 95258 | 31 AA2: 32 | MKRFLIYLKSRPLGMASVIALIVLYLMMVFA |
| 95283 | Probability: | MKTKKIKSKAEMVVIACAVIAAVAVLNYA |

| | | |
|---|---|---|
| | 0.998 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| | 0.976 AA1: | MDFIFKRPGSKFNIRLLAYFAVIVIFTIAPI |
| 95288 | 40 AA2: 41 | TLMTSAALN |
| | Probability: | |
| | 1.000 AA1: | |
| 95341 | 26 AA2: 27 | MRTQFSKIAFLAGFALAMSFPLSCSG |
| | Probability: | |
| | 1.000 AA1: | |
| 95347 | 22 AA2: 23 | VSVRRKFLILAILCLISFPAFS |
| | Probability: | |
| | 0.994 AA1: | |
| 95349 | 19 AA2: 20 | MKINSALIAGAGAIGLCVA |
| | Probability: | |
| | 0.997 AA1: | |
| 95358 | 27 AA2: 28 | MRTVHFAKIAGIVLLPVFLLSCFMAAG |
| | Probability: | |
| | 0.504 AA1: | |
| 95360 | 17 AA2: 18 | IFPAILLIVTTHPSEFA |
| | Probability: | |
| | 0.778 AA1: | |
| 95380 | 13 AA2: 14 | MAIFLTALGAALA |
| | Probability: | |
| | 1.000 AA1: | |
| 95401 | 24 AA2: 25 | MWFLTAAVLCAVSVTVGCSAKAEA |
| | Probability: | |
| | 0.999 AA1: | |
| 95405 | 20 AA2: 21 | MKIKFKLSIVVIAIMAAVTA |
| | Probability: | |
| | 0.963 AA1: | |
| 95410 | 25 AA2: 26 | MKKVYIGVAAACVLLIVILATGLWK |
| | Probability: | |
| | 0.988 AA1: | |
| 95420 | 29 AA2: 30 | MKIRSTAVPVILFAALIILLTLAPFVILN |
| | Probability: | |
| | 0.959 AA1: | |
| 95449 | 23 AA2: 24 | LISQILLSLVILAVAISIGTVTT |
| | Probability: | |
| | 0.899 AA1: | |
| 95451 | 20 AA2: 21 | MIKRYLLLYFFVLTIPLCLA |
| | Probability: | |
| | 0.513 AA1: | |
| 95457 | 21 AA2: 22 | MDTITVCMGSSCFGKGNSANA |
| | Probability: | |
| | 0.993 AA1: | |
| 95459 | 19 AA2: 20 | MKIKIKLSLIVIAIVVVVA |
| | Probability: | |
| | 0.999 AA1: | |
| 95464 | 26 AA2: 27 | MKSSPSVKVFLITLGLLLYAPAVSQA |
| 95469 | Probability: | MIRFVIVFVIILLALAVVITLLVA |

| | | |
|---|---|---|
| | 0.994 AA1: 24 AA2: 25 Probability: 1.000 AA1: | |
| 95473 | 19 AA2: 20 Probability: 0.911 AA1: | MMKPLFISVFILVSAGLFA |
| 95480 | 25 AA2: 26 Probability: 0.999 AA1: | MRRNLFLKLIVTISLTTLLANCGLG |
| 95481 | 18 AA2: 19 Probability: 1.000 AA1: | LKTGATASLTVTVSPANA |
| 95515 | 17 AA2: 18 Probability: 0.672 AA1: | MKRLFLLLLLIPSFVFA |
| 95517 | 25 AA2: 26 Probability: 1.000 AA1: | MRKLYNYPWLIVAVIAVITVFFSFQ |
| 95518 | 21 AA2: 22 Probability: 0.998 AA1: | MKKNKFIALAFLFFVTVPASS |
| 95526 | 15 AA2: 16 Probability: 0.891 AA1: | MFIFYLLALAAIALA |
| 95549 | 40 AA2: 41 Probability: 0.574 AA1: | MQTERSVSQPVVGRTPLPMRFVAVSFLMVFICAGFAAVFA<br>MGSSGSSWSVMIAATESYILKEVKFLTKITIIIAAVAIVIA |
| 95555 | 53 AA2: 54 Probability: 1.000 AA1: | AIIVFLTLSATT |
| 95564 | 20 AA2: 21 Probability: 0.999 AA1: | MKIKLIAVFMGFVVLTGAFA |
| 95565 | 46 AA2: 47 Probability: 0.976 AA1: | LFQKAGLRLSRAINRKHTDIFFKYLFSTVALFSVLALGAILFFVFA |
| 95571 | 31 AA2: 32 Probability: 0.940 AA1: | DKAAYKLTWIFLILLFAFFGGIMYIFFYSQS |
| 95572 | 23 AA2: 24 Probability: 0.989 AA1: | VNWFSSNTAVATVNGGTVTAVSA |
| 95579 | 25 AA2: 26 Probability: 1.000 AA1: | MNGIGKFSLTAGIGGLLISCMAGTA |
| 95601 | 19 AA2: 20 Probability: 0.974 AA1: | MKKLLVIMLVLLVASMAFA |
| 95610 | 25 AA2: 26 | MKKISTLLIGICLLLTVIGCQNPFQ |
| 95620 | Probability: | MLKAGKKSTAALSVLSAFIIILVSASAVR |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 29 AA2: 30 | |
| | Probability: | |
| | 1.000 AA1: | |
| 95634 | 21 AA2: 22 | MKSRLLFTLCLLFTLTAAIYG |
| | Probability: | |
| | 0.914 AA1: | |
| 95636 | 41 AA2: 42 | MKRNVNIFMINSRKRIIPFFQYFRSGPVLFSFVLFIIAAFV |
| | Probability: | |
| | 0.715 AA1: | |
| 95644 | 20 AA2: 21 | MKKIIERLLVFIIGIPAVYA |
| | Probability: | |
| | 1.000 AA1: | |
| 95664 | 18 AA2: 19 | VKKKLLGFLLLISLPVFS |
| | Probability: | |
| | 0.999 AA1: | |
| 95687 | 31 AA2: 32 | MKLIKKLLGITAIIAAVGFITLPLTGCPAES |
| | Probability: | |
| | 0.862 AA1: | |
| 95702 | 22 AA2: 23 | IESLVYALGVALGFMLSLLLLA |
| | Probability: | |
| | 0.975 AA1: | |
| 95742 | 24 AA2: 25 | MKKLNSRFLLFFIPALVLHTACSG |
| | Probability: | |
| | 0.988 AA1: | |
| 95750 | 28 AA2: 29 | MKINSGFTTKYLSVIAATLLFFIAFVTG |
| | Probability: | |
| | 0.999 AA1: | |
| 95753 | 18 AA2: 19 | MKFKIKMSCMMIAIVAIA |
| | Probability: | |
| | 1.000 AA1: | |
| 95755 | 37 AA2: 38 | MISNRKNYRNAFAAASLLLMLTMTAAASTSITASVSA |
| | Probability: | |
| | 0.735 AA1: | MFNIFSEMFSYAFLVRAFIVGALVSICAALL |
| 95757 | 63 AA2: 64 | GVSLVLKRYSMIGDGLSHVGFGSLAVATALNA |
| | Probability: | |
| | 0.968 AA1: | MIAVLIKMFSYSFLIKAFTVGALVSICAALLGVS |
| 95760 | 63 AA2: 64 | LVLKRYSMIGDGLSHVGFGSLAVATALHA |
| | Probability: | |
| | 0.997 AA1: | |
| 95768 | 25 AA2: 26 | MKRVFWSIVLFTGICAMLFTQNTGA |
| | Probability: | |
| | 0.711 AA1: | |
| 95774 | 19 AA2: 20 | MDKKALIAMSGGVDSSVAA |
| | Probability: | |
| | 0.841 AA1: | |
| 95786 | 33 AA2: 34 | MTISEYSVKRPVTVVILFALCIGIAFTMLVNLA |
| | Probability: | |
| | 1.000 AA1: | |
| 95802 | 19 AA2: 20 | MKFRVFLFVFIFACVFTNA |
| 95817 | Probability: | LSFLKFQSKNQYADINDMKKNYL |

| | | |
|---|---|---|
| | 0.999 AA1: | FLAILFLSVIIPFQLQA |
| | 40 AA2: 41 | |
| | Probability: | |
| | 0.950 AA1: | |
| 95837 | 23 AA2: 24 | MAAYIRRSTILFALPLVAVVLFA |
| | Probability: | |
| | 0.999 AA1: | |
| 95844 | 31 AA2: 32 | MSLRIKLTTIILVMILVSVAVLSVFNLSRSA |
| | Probability: | |
| | 0.952 AA1: | |
| 95852 | 36 AA2: 37 | MAKNKNPMRSEFIIFSSVLFLVILAVGGLAFFFSMR |
| | Probability: | |
| | 1.000 AA1: | |
| 95865 | 23 AA2: 24 | MRRFGIIICLFFILSFFPAALSA |
| | Probability: | |
| | 0.962 AA1: | |
| 95870 | 30 AA2: 31 | MKSSKLYLIVAASVVLLAVGGLLFINNLSN |
| | Probability: | |
| | 1.000 AA1: | |
| 95876 | 31 AA2: 32 | MKSKKIAALFLTAAVLCAVLVTAGCSVRAEA |
| | Probability: | |
| | 0.801 AA1: | MDMILSESFITPVGIYRQALIAFVFVALFFLASM |
| 95913 | 52 AA2: 53 | HFSPPARVSSGFILAALA |
| | Probability: | |
| | 1.000 AA1: | |
| 95946 | 23 AA2: 24 | MKRNILPWVVLLLFLSSCNTQSA |
| | Probability: | |
| | 0.516 AA1: | |
| 95960 | 36 AA2: 37 | MINKYRLSIIDLHKNVYLRGLCLIIFLLFSVKNIYA |
| | Probability: | |
| | 1.000 AA1: | |
| 95968 | 22 AA2: 23 | MLHMKKLSLLLLGSSFAVMLSA |
| | Probability: | |
| | 0.917 AA1: | |
| 95998 | 41 AA2: 42 | GSSSEVFAVFGKNIERIKKNLLLLAISGGIVAIATTLLLVA |
| | Probability: | |
| | 1.000 AA1: | |
| 96009 | 28 AA2: 29 | MKKRLIFTFIVIASVLLFLAASPWEGSA |
| | Probability: | |
| | 1.000 AA1: | |
| 96015 | 24 AA2: 25 | MKFNRLHILCLISLFTSVCFNIYA |
| | Probability: | |
| | 0.802 AA1: | |
| 96020 | 44 AA2: 45 | MRKIIYFIIEAFRGFYQAKLMTFVSIMTIAATLFLICGAFIALM |
| | Probability: | |
| | 1.000 AA1: | |
| 96024 | 22 AA2: 23 | MKLNAKLLVFGACLLMTRPLFA |
| | Probability: | |
| | 1.000 AA1: | |
| 96033 | 24 AA2: 25 | MKILKKLLIIVPITILLIFLVLLA |
| 96039 | Probability: | MIGLFALAFFLQILAFVFS |

| | | |
|---|---|---|
| | 0.868 AA1: 19 AA2: 20 Probability: | |
| | 0.793 AA1: | |
| 96041 | 26 AA2: 27 | MKTRRFCIIYFAVTLLFLFYGCKSVS |
| | Probability: 0.845 AA1: | |
| 96059 | 34 AA2: 35 | LNTRKLILSLAVNWPAKVLSVALALILFVFHEMS |
| | Probability: 0.729 AA1: | |
| 96071 | 27 AA2: 28 | NFKRRFIVIFSLLAIFLLYLIIHYAVL |
| | Probability: 0.718 AA1: | |
| 96076 | 17 AA2: 18 | MNGLFLLVISVVVLSLA |
| | Probability: 0.999 AA1: | |
| 96082 | 29 AA2: 30 | MKKPKKTSSAMLFITVSLLMILSITVSLS |
| | Probability: 0.563 AA1: | |
| 96088 | 22 AA2: 23 | MSISNLTWGILTCSWFGFDVSA |
| | Probability: 1.000 AA1: | |
| 96091 | 24 AA2: 25 | MTSLKVRKALSIAVLLATSSTVFA |
| | Probability: 0.886 AA1: | |
| 96094 | 33 AA2: 34 | KGFIALAAVSFGRWLPLGVLGASFLFGFSSSLA |
| | Probability: 0.977 AA1: | |
| 96099 | 20 AA2: 21 | MIMSDILFLFFLGLLIVALA |
| | Probability: 0.752 AA1: | |
| 96113 | 27 AA2: 28 | IRVAAYVSIAMLIAVIIVAVIMSLSIS |
| | Probability: 0.983 AA1: | |
| 96119 | 37 AA2: 38 | MKKIIPRLLVFFIGLPIIFAMVFLLPQCRHLVFNLFA |
| | Probability: 0.876 AA1: | |
| 96140 | 20 AA2: 21 | MKVFVIAIILSLLFVGLLVG |
| | Probability: 0.748 AA1: | |
| 96155 | 28 AA2: 29 | MNFKRRFIVIFSLLAIFLLYLIIHYAVL |
| | Probability: 0.841 AA1: | |
| 96161 | 32 AA2: 33 | MLAPRILQYMPRKFNRLLEPFAGMAALTLAAA |
| | Probability: 0.968 AA1: | |
| 96179 | 30 AA2: 31 | MSSKQKIIMISAGAALAVLIVLGIIFFVKG |
| | Probability: 1.000 AA1: | |
| 96184 | 35 AA2: 36 | MTKKNAAFLLALVPVLAAAFSFILPYAASFTPAVA |
| 96208 | Probability: | MKRTVFLSLALFALMAVGVYA |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 0.736 AA1: | |
| 96212 | 23 AA2: 24 | MKKLLRLLGIIVITAIIGFSIIT |
| | Probability: | |
| | 0.997 AA1: | |
| 96216 | 25 AA2: 26 | MKTFFYSLLPIVLLFAFMSCSNNNS |
| | Probability: | |
| | 0.889 AA1: | |
| 96225 | 32 AA2: 33 | MKKSILILITLAVFLFGISGCKKNKNAVTRNA |
| | Probability: | |
| | 0.778 AA1: | |
| 96241 | 19 AA2: 20 | MKHVITSIMLSAMPFFIFS |
| | Probability: | |
| | 1.000 AA1: | |
| 96256 | 22 AA2: 23 | MKLNAKLLVLGAYLLMARPLFA |
| | Probability: | |
| | 0.718 AA1: | |
| 96262 | 23 AA2: 24 | FTVILVAVAIVVTAAIVYFVLHS |
| | Probability: | |
| | 0.994 AA1: | |
| 96264 | 15 AA2: 16 | MWVILLLFSVLAVFA |
| | Probability: | |
| | 0.953 AA1: | |
| 96269 | 19 AA2: 20 | MAVMLFLLGLAVFGLMALA |
| | Probability: | |
| | 0.992 AA1: | LRILIKYLKPFTLLLFLAVMILFLQ |
| 96274 | 49 AA2: 50 | VFCDLSLPNLMSRLVNSILAQSAA |
| | Probability: | |
| | 0.585 AA1: | |
| 96285 | 16 AA2: 17 | MRKSSFLALCMNPTLQ |
| | Probability: | |
| | 0.983 AA1: | |
| 96294 | 24 AA2: 25 | MRLTPVFTIFLLLLFFACSGEKNT |
| | Probability: | |
| | 0.980 AA1: | |
| 96304 | 21 AA2: 22 | MKKIIILPLLFLSLVSCEAMK |
| | Probability: | |
| | 0.801 AA1: | |
| 96328 | 35 AA2: 36 | MNLNGIWQLKKIAMRNLARHKVKTLLTSAAIMVSV |
| | Probability: | |
| | 0.992 AA1: | |
| 96355 | 37 AA2: 38 | MKRSRSKSLNQEMNLTNMIDVVFAILIVFLISAPLMS |
| | Probability: | |
| | 0.996 AA1: | |
| 96367 | 25 AA2: 26 | MFSSVDTMWVLVAAALVFFMQAGFA |
| | Probability: | |
| | 0.988 AA1: | |
| 96380 | 40 AA2: 41 | MRIGGRNPAFFVHTKAGGRLPLPQFLFMLLVLMFVPSAFA |
| 96381 | Probability: | MNAPITRTLLTTAVALSLALAGAGCG |

| | | |
|---|---|---|
| | 0.991 AA1: 26 AA2: 27 Probability: 1.000 AA1: | |
| 96382 | 29 AA2: 30 Probability: 0.939 AA1: | MNTFFHSAKTTTAIIWVAVLLFLSFAATA |
| 96424 | 25 AA2: 26 Probability: 0.995 AA1: | VIKILKYLKPKEWLMAALSLIFIVA |
| 96427 | 21 AA2: 22 Probability: 0.506 AA1: | MNHRPFAIILAFLFLPAVIFA |
| 96435 | 23 AA2: 24 Probability: 0.695 AA1: | VERALRVLDGAILVLCSVGGVQS |
| 96457 | 32 AA2: 33 Probability: 0.763 AA1: | MAIIPILLRLAGGLCMFLFGMKMMSDGIQQSA |
| 96458 | 35 AA2: 36 Probability: 0.999 AA1: | AKFFANRSRAKAKAAATAQTTVEEDSASISTAAAA |
| 96461 | 17 AA2: 18 Probability: 0.786 AA1: | MFRFLIFILISAIFAFA<br>MIGKKIGFINDKIYGIAMPIGFTVIAIRFILKLQNTRD |
| 96462 | 58 AA2: 59 Probability: 0.885 AA1: | KILASAVFLIGCIAALPAIA |
| 96468 | 23 AA2: 24 Probability: 1.000 AA1: | MKLSIDTLFFLVPVGSILALLFA |
| 96481 | 22 AA2: 23 Probability: 0.563 AA1: | MFLRKIILIISLIFFISAALWA |
| 96491 | 26 AA2: 27 Probability: 1.000 AA1: | MAQKEYSVGYALFIACAAAIGGFLFG |
| 96502 | 31 AA2: 32 Probability: 0.918 AA1: | MRNDAFFTSAKKTAAIISAAVVLCFVFAASA |
| 96512 | 21 AA2: 22 Probability: 0.977 AA1: | MKTNKSGWALVLSGGGAMGLA |
| 96514 | 18 AA2: 19 Probability: 0.506 AA1: | LALCVVAMAFAGGSQASG |
| 96517 | 26 AA2: 27 Probability: 0.901 AA1: | MKIKNSGASWIPWLAPALVFYVFFMA |
| 96526 | 15 AA2: 16 | MLIGCLGVLLAISAA |
| 96537 | Probability: | MSCRRYPRRAVTAAMTVMLTAFALYGLA |

| | | |
|---|---|---|
| | 1.000 AA1: 28 AA2: 29 Probability: 0.653 AA1: | |
| 96547 | 35 AA2: 36 Probability: 0.903 AA1: | LKLRNILEIDYLLLLPVIILMMLGVFFIYSSGITS |
| 96578 | 19 AA2: 20 Probability: 0.697 AA1: | MIYTVLIFVGAAIAGALVS |
| 96583 | 30 AA2: 31 Probability: 0.994 AA1: | MIKLSAIFIAFFRVGAFTLGGGYAMFSVLR<br>AYNVNVNIKVNIKTLDCFASLAMTVIAFT |
| 96591 | 62 AA2: 63 Probability: 0.986 AA1: | TTVIAFTTTVIAFTTTVIAFTTTVIAFTTTVIA |
| 96674 | 35 AA2: 36 Probability: 0.994 AA1: | MNLNKLRLKWKIFAFLLGFCALLLIVLWLFQIVFL |
| 96675 | 28 AA2: 29 Probability: 0.517 AA1: | MLPVCFSFLCWFSSTALMAIAGYLIARA |
| 96687 | 18 AA2: 19 Probability: 0.607 AA1: | MTAQTAVIGAGAWGTAVA |
| 96726 | 20 AA2: 21 Probability: 0.902 AA1: | LKHIGVLGIMTSCSILDAVG |
| 96728 | 37 AA2: 38 Probability: 1.000 AA1: | MVIMLFLGFKVNRRTGHALIAAIGFAALSLTGLSLTG |
| 96760 | 23 AA2: 24 Probability: 0.806 AA1: | MKISFKLMSIMIALGLFAVAAVS |
| 96766 | 20 AA2: 21 Probability: 0.998 AA1: | MKILLPILTIFTCLGFAGHG |
| 96780 | 20 AA2: 21 Probability: 0.651 AA1: | MKIQFKLSIVVIAIMAAVVA |
| 96807 | 21 AA2: 22 Probability: 0.970 AA1: | MSKRYRFAIIIVIVAVCFVFL |
| 96820 | 38 AA2: 39 Probability: 0.999 AA1: | IFLRNIKGTLIIGLARPISILLTLMCMAFFGFTLNLLT |
| 96830 | 30 AA2: 31 Probability: 0.931 AA1: | MNLKYRYAGSLQIPLFLAIFLLLCAYYSHA |
| 96846 | 20 AA2: 21 | STICVIAFVAIIGFSFPACG |
| 96850 | Probability: | MNGILVLLLLLLTATLPAIVVFFWLRSGK |

| | | |
|---|---|---|
| | 0.985 AA1: 51 AA2: 52 Probability: 1.000 AA1: | SPVTLPLFLLSLAMGIISFVVA |
| 96851 | 18 AA2: 19 Probability: 0.818 AA1: | AAAAPAAAAPAAAAPAAA |
| 96854 | 29 AA2: 30 Probability: 0.997 AA1: | MNKRFFHFIFCSLIFTVSAIILFPTEDIA |
| 96861 | 25 AA2: 26 Probability: 0.941 AA1: | MKMNRKLFIALATLLLFGMLAGCRK |
| 96863 | 27 AA2: 28 Probability: 0.851 AA1: | MKFLHALEMAGIFLLFAAAAGPHFVFT |
| 96870 | 29 AA2: 30 Probability: 0.673 AA1: | VSRRLDKSPNRILAKAAAPGPSIPVQSAA |
| 96875 | 22 AA2: 23 Probability: 1.000 AA1: | GLSLATVITAIIVVSCASTGNK |
| 96876 | 31 AA2: 32 Probability: 0.635 AA1: | VKFIRFAANSFKFCLFALLLAGLGLAAPLYG TSIGGGSYSRSFIGCSALTSITIPSSVTFIGTDSFS |
| 96887 | 47 AA2: 48 Probability: 0.980 AA1: | GCTALTSITVA |
| 96904 | 23 AA2: 24 Probability: 0.998 AA1: | GNGNTGSSAAGASSGTTTTGSSA |
| 96910 | 31 AA2: 32 Probability: 0.999 AA1: | MLRKFRIRTRLLFSFFIVVLFTLIVGLTGYA |
| 96914 | 19 AA2: 20 Probability: 0.892 AA1: | MKKNFICILLLLGYVSLYA |
| 96915 | 25 AA2: 26 Probability: 0.998 AA1: | DVSFSSLFINRPVATTLLIAAVALG |
| 96916 | 18 AA2: 19 Probability: 0.549 AA1: | MKTRALFLIFLFSLAICA VSNPSSGNFAGIWVFNGNVNVGTLLAGQA |
| 96932 | 39 AA2: 40 Probability: 0.996 AA1: | VGAAAPNATA |
| 96941 | 25 AA2: 26 Probability: 0.999 AA1: | MKNTRKFINFAVIAAIGFVVLSLTG |
| 96942 | 34 AA2: 35 | MKIKFKLSILIIAIVAVIVSGVAVLLLRQASSIA |
| 96944 | Probability: | VAVIVTGVAALLLRQASSIA |

| | | |
|---|---|---|
| | 0.801 AA1: 20 AA2: 21 Probability: 0.999 AA1: | |
| 96950 | 27 AA2: 28 Probability: 1.000 AA1: | MKMKLTAAIIAAFVIIMGCAGGPKASK |
| 96962 | 23 AA2: 24 Probability: 0.951 AA1: | MSNIGKITAFVIFICLFPLPVFA |
| 96965 | 21 AA2: 22 Probability: 0.995 AA1: | LKGILLAGGAGSRLYPLSAVA |
| 96969 | 30 AA2: 31 Probability: 0.964 AA1: | MGGNMREQIIIRKAAFFLIVLFLLSTGLYA |
| 96970 | 23 AA2: 24 Probability: 0.808 AA1: | MKKYLKQAVILFILLFICISCSG |
| 96978 | 16 AA2: 17 Probability: 0.998 AA1: | MKNFTGIPVSAGVAYA MSDNGSPDKMPSRRRMPFSWIWFASTL |
| 96988 | 43 AA2: 44 Probability: 0.653 AA1: | VLCAVFAFASVPSVLA |
| 96992 | 19 AA2: 20 Probability: 0.643 AA1: | MRNIQKLLPLLMIVLPLVS MTKTMQYRSVALFNKSEKDLIIYTGNKEN |
| 97004 | 62 AA2: 63 Probability: 0.981 AA1: | KIRHIALIKKISSITIICVLLAFIVISIFSIHS |
| 97008 | 29 AA2: 30 Probability: 0.531 AA1: | MESKKKSNGKFRFVDFLVILLCLSGAVYS |
| 97010 | 37 AA2: 38 Probability: 0.978 AA1: | MKKRNSKNRMVTPQANGSTMKEPLPPALLSAPGQRPA |
| 97021 | 22 AA2: 23 Probability: 0.532 AA1: | MKIAVKGLGIAILAVMLVFALA |
| 97028 | 19 AA2: 20 Probability: 0.659 AA1: | EAVSVSVGSVTISGGKVSA |
| 97029 | 30 AA2: 31 Probability: 0.691 AA1: | MSDTANKERRKTSNKALDIFVVLFCLCGAA |
| 97031 | 33 AA2: 34 Probability: 1.000 AA1: | LLGVAANFFTVAGATAGNPANSGVVTAVFPATA |
| 97043 | 23 AA2: 24 | MKRKLSIMAAMAAILICCLVCTT |
| 97046 | Probability: | MKNILTVLIITVIVFTFFTCSTSSNG |

| | | |
|---|---|---|
| | 1.000 AA1: 26 AA2: 27 Probability: 0.976 AA1: | |
| 97093 | 21 AA2: 22 Probability: 1.000 AA1: | MRNMLAVLHSAVTIMTPLLFA |
| 97124 | 36 AA2: 37 Probability: 0.997 AA1: | MLNNKKDPVFIVKVCMLMLVGAFVFSSVSAAFNARA |
| 97133 | 23 AA2: 24 Probability: 0.633 AA1: | MKIKIKLSIMMIAIVAIVAGGVA |
| 97134 | 33 AA2: 34 Probability: 0.961 AA1: | MRFSYFRDRRHTAVPAIIIFCVAILYISCAGNT<br>MAKTSQRKSFIKVTLASFAVSSLLFLTSLDN |
| 97141 | 45 AA2: 46 Probability: 1.000 AA1: | KFFDLFLRFLPSLT |
| 97165 | 17 AA2: 18 Probability: 0.941 AA1: | VTKKTAACLLALPPVLA |
| 97169 | 19 AA2: 20 Probability: 0.993 AA1: | IRKYIVLAVVFAFCFFVYM |
| 97181 | 22 AA2: 23 Probability: 0.997 AA1: | LKTIKYLFFALFVFFSIIFLNA<br>MICMNRKITDVFFKHLFSIV |
| 97224 | 48 AA2: 49 Probability: 1.000 AA1: | ALFSVLALGAILLFVFVQGGMPFFSSTA |
| 97246 | 23 AA2: 24 Probability: 1.000 AA1: | MKICVYAAALSALSIMIASCANQ |
| 97254 | 27 AA2: 28 Probability: 0.869 AA1: | MKYIYRLLAAICSIFCTVCSCATGTEA |
| 97258 | 17 AA2: 18 Probability: 1.000 AA1: | MALQKAIWLLAAVALNA |
| 97259 | 22 AA2: 23 Probability: 1.000 AA1: | MVKKIAFLFLLFVIAVSSASFA |
| 97272 | 32 AA2: 33 Probability: 1.000 AA1: | MNNRNRIIAGVIISLLIIIAAASLGSSGISLR |
| 97273 | 19 AA2: 20 Probability: 0.765 AA1: | MKRVVFMFAFILAGISAFA |
| 97282 | 32 AA2: 33 | MIDTIINMFSFAFLVRAFVVGALVSICASLLG |
| 97286 | Probability: | MRINHILTTRAKATVAALFFIAAISHPALS |

| | | |
|---|---|---|
| | 1.000 AA1: 30 AA2: 31 Probability: 0.738 AA1: | |
| 97290 | 31 AA2: 32 Probability: 0.998 AA1: | MENKKYLKFKEKYGPYALVAGGSNGLGAAYA |
| 97309 | 34 AA2: 35 Probability: 0.976 AA1: | MSLKVKLTTIILVMILVSIAVLSVFNLTRSSALQ |
| 97320 | 35 AA2: 36 Probability: 0.996 AA1: | MVSAIHRVAGKRFAGLSLAVAALVLLPLFPGRVLR |
| 97323 | 32 AA2: 33 Probability: 0.859 AA1: | VRTETFDVVILGGGAAGLSAALAAAEAGVAVA |
| 97336 | 27 AA2: 28 Probability: 0.999 AA1: | MHQLPQLSAIHRLVWIALMAALTAAGA VLIGNNISNIASSTLATALFIGLFGVNGVSIATLIMT |
| 97337 | 52 AA2: 53 Probability: 0.993 AA1: | ILLMLFCDISPKALA |
| 97348 | 22 AA2: 23 Probability: 0.996 AA1: | VQSIWKRALIAFAGPAFNIVFA |
| 97352 | 26 AA2: 27 Probability: 0.964 AA1: | MITKLIKIALMVSIALSVSFTLSCGQ |
| 97373 | 27 AA2: 28 Probability: 1.000 AA1: | MAILSTVLLVFFVIIAILLVLLVLVQN |
| 97375 | 23 AA2: 24 Probability: 1.000 AA1: | MNKKHLLAFAFGLMLTFSGYAQS |
| 97379 | 20 AA2: 21 Probability: 1.000 AA1: | MKKVITSLLISAALATVAFA |
| 97382 | 22 AA2: 23 Probability: 0.612 AA1: | MKRSTLLCILSGVFLAATLAPA |
| 97383 | 16 AA2: 17 Probability: 0.997 AA1: | GSHTISLTVPTVVAIA |
| 97399 | 19 AA2: 20 Probability: 1.000 AA1: | MPWRICVALAVFALSLVYA |
| 97429 | 21 AA2: 22 Probability: 0.974 AA1: | MKRSVFCIIALFGFLSVSAFS MKVKKKRKNQAQAVLAAPAVSDNKYLALSFI |
| 97486 | 42 AA2: 43 | FPVAIMGTVFA |
| 97493 | Probability: | MKKILLPILILLSLGLQA |

| | | |
|---|---|---|
| | 1.000 AA1: 18 AA2: 19 Probability: 0.962 AA1: | |
| 97498 | 25 AA2: 26 Probability: 0.996 AA1: | MKSILIKGYVFVAVAALIVSVYFIA |
| 97499 | 41 AA2: 42 Probability: 1.000 AA1: | MRETRIKPVKIKKRRLKYIAVLPSFITLINGACGFIAIVFA |
| 97512 | 20 AA2: 21 Probability: 1.000 AA1: | MKKIISVCFLGFFITASAFA |
| 97515 | 29 AA2: 30 Probability: 0.970 AA1: | MLNRYQRAFALLILTAILTAILTATPAPA |
| 97524 | 25 AA2: 26 Probability: 0.909 AA1: | MYLKYVFGALIILSGLLLHSCSSYS |
| 97529 | 22 AA2: 23 Probability: 0.988 AA1: | MKTRELVILGAGPAGLTAAVYA |
| 97534 | 19 AA2: 20 Probability: 0.983 AA1: | MKIKYKLSVMMIAIVLVVA |
| 97537 | 35 AA2: 36 Probability: 0.927 AA1: | MNLNKLRLKWKIFAFLLGFSALLLIVLWLFQTILL |
| 97542 | 19 AA2: 20 Probability: 1.000 AA1: | VRHAALPLIISIGLCSCAA |
| 97553 | 22 AA2: 23 Probability: 0.999 AA1: | MKKIFKVLAIIAIVAVIGFSMA |
| 97555 | 31 AA2: 32 Probability: 1.000 AA1: | MKTILKKNYIFVAAAILIVSAYFISGVLVFA |
| 97570 | 21 AA2: 22 Probability: 1.000 AA1: | MKKILALAVALFIFVPAAVSA |
| 97587 | 26 AA2: 27 Probability: 0.665 AA1: | MNKTVRTIVWAVMTAVLTVGAAGAWA |
| 97623 | 31 AA2: 32 Probability: 0.990 AA1: | RDKTQIVGKKRFVIFASVLALASLGILGRYG |
| 97627 | 32 AA2: 33 Probability: 0.726 AA1: | MTLGAVRFAIAVAALFLMYRFSAKRGAAAAQA |
| 97644 | 16 AA2: 17 | GRTGSALSGLPGNASS |
| 97647 | Probability: | MKKLVVALAVVVLLLPGRPALS |

| | | |
|---|---|---|
| | 1.000 AA1: 22 AA2: 23 Probability: 0.997 AA1: | |
| 97687 | 33 AA2: 34 Probability: 0.585 AA1: | MRTYHSLPALVIAVMLAGLITLSCASTGGGRKA |
| 97725 | 28 AA2: 29 Probability: 0.864 AA1: | MIIFEFVLIMLAAVLLSNLINRYIPALS |
| 97729 | 19 AA2: 20 Probability: 1.000 AA1: | MKKIFLILITVAAVFYGSG |
| 97745 | 20 AA2: 21 Probability: 0.981 AA1: | MRKLILTTAAVAAAIAGAFA |
| 97747 | 23 AA2: 24 Probability: 1.000 AA1: | LLIILLPSIFPCSLFALSFSDSG |
| 97752 | 20 AA2: 21 Probability: 0.794 AA1: | MKHFTTLLFLAAALCATSCA |
| 97759 | 31 AA2: 32 Probability: 0.785 AA1: | VGLPVMERELGLSAVMLGWIPLSLMMPMAAA |
| 97770 | 50 AA2: 51 Probability: 0.743 AA1: | MPRNKEEGIRNKIKMRRGFAATVNTKRALRVLTAFLIPY SLFLIPCALHS |
| 97771 | 65 AA2: 66 Probability: 0.961 AA1: | MVITLFSGELLKLSSKAKLLNISLLSVAVMYFMMACSIRP ATFIALDFLSAVPLMLVSSLLSLFM |
| 97780 | 18 AA2: 19 Probability: 0.598 AA1: | IVTLLAGFSAAAFLSVFK |
| 97797 | 18 AA2: 19 Probability: 0.807 AA1: | IKRALLVAAAGGHNILMV |
| 97868 | 36 AA2: 37 Probability: 0.819 AA1: | MIIKKNFSDYLIDGFIYAFMILLIVLVLYPFFNALA |
| 97872 | 25 AA2: 26 Probability: 0.795 AA1: | MKFDVILAGVGGQGVLSVAAIIAQA |
| 97873 | 38 AA2: 39 Probability: 0.903 AA1: | ALFTGLFFLVALFFWPVFRIVPIAATFPALIFAGFLMM |
| 97887 | 40 AA2: 41 Probability: 0.770 AA1: | MKKKRKGLTMRQREIVTGFLFISPWIVGVAYFFIRSIAQA |
| 97890 | 35 AA2: 36 | MDLSVLIAIGAGIAVATGIGAGIGIGIATSGLVQA |
| 97896 | Probability: | MTRFKLYLLLGAFCLVLGCRAKYG |

| | | |
|---|---|---|
| | 0.975 AA1: 24 AA2: 25 Probability: 0.980 AA1: | |
| 97910 | 40 AA2: 41 Probability: 1.000 AA1: | MNKNKKSGRLGLQLKLSSISITFMLLAIVIFSIISIRAIQ |
| 97911 | 20 AA2: 21 Probability: 0.507 AA1: | VVRAVLLALLGIIIPLYAGA |
| 97920 | 21 AA2: 22 Probability: 0.885 AA1: | MKITRYLIAILPLALCLYCDK |
| 97933 | 20 AA2: 21 Probability: 0.967 AA1: | FEAILCGGMIFMAMVTFVLA |
| 97964 | 24 AA2: 25 Probability: 0.572 AA1: | MKKIVLIQVLAVAVLATCCTTAGK |
| 97973 | 36 AA2: 37 Probability: 0.549 AA1: | MKSIPWQIKLIASILLIILILWKFPVSEIIANLQNA |
| 97981 | 40 AA2: 41 Probability: 0.625 AA1: | VSLGEVTMLLRGGIWQVLTLAAPLLLSALAVGLIVAILQA |
| 97989 | 36 AA2: 37 Probability: 0.986 AA1: | MGKNQQDKVRFSIGNKLIIIISIIVILSLGSITALV |
| 97998 | 31 AA2: 32 Probability: 0.982 AA1: | VRGKNKKWFVFSIVLFMFAVTNLIIPIVIFA |
| 98018 | 26 AA2: 27 Probability: 0.998 AA1: | MRPLIHALTKFFAALLVIALPFVSCS VRSLKFKNRTPKRRSIYILPNIFTTLSLFFGFYS |
| 98037 | 46 AA2: 47 Probability: 1.000 AA1: | ILAAIEARYSLA |
| 98064 | 19 AA2: 20 Probability: 0.989 AA1: | MKKTMYAILSLFAITVLFA |
| 98132 | 29 AA2: 30 Probability: 1.000 AA1: | MRKAIRNNVGMIRTLTACVSVIITLAILA |
| 98133 | 30 AA2: 31 Probability: 0.999 AA1: | LKPSPVRKFAAFLPFAVLILISAATVKTTA |
| 98137 | 25 AA2: 26 Probability: 1.000 AA1: | MKSFKIPTFVLVLLLCLCFAMAGCG |
| 98147 | 20 AA2: 21 | MKKKLLICGLFLLAAFTVFS |
| 98165 | Probability: | MPASRDSTSQARQPVSRRQVIQAAAGISAGL |

| | | |
|---|---|---|
| | 0.804 AA1: 41 AA2: 42 Probability: 0.999 AA1: | ALSQVSSRAA |
| 98175 | 25 AA2: 26 Probability: 0.756 AA1: | MSARAFCIFLLGFLALTVCAHIALA |
| 98180 | 23 AA2: 24 Probability: 0.993 AA1: | MWIGVLLAAALCLGLGIFINETT |
| 98185 | 23 AA2: 24 Probability: 1.000 AA1: | MKHHNLNSFLTALFFSLLFTVCA |
| | | MESSKLPNSKSARGSRLAGAVVAA |
| 98204 | 39 AA2: 40 Probability: 0.890 AA1: | SVLAAAFVSPRSAQA |
| 98209 | 19 AA2: 20 Probability: 0.998 AA1: | LRLVRLLGALDLPVPPASA |
| 98249 | 25 AA2: 26 Probability: 0.993 AA1: | MKDPIRFRFAGIFMILAMCAVFGLA |
| 98258 | 31 AA2: 32 Probability: 0.593 AA1: | MNNETKVGMAVIAGLVVLMGMVSFLGLFRFA |
| 98260 | 21 AA2: 22 Probability: 0.824 AA1: | MRQSILAVIGIILFLVFSRLN |
| 98316 | 26 AA2: 27 Probability: 0.784 AA1: | MALTLSLLGIALASVLIIVLIRVDAA |
| 98340 | 19 AA2: 20 Probability: 0.987 AA1: | ADSISASPSISFSRSAARA |
| 98342 | 29 AA2: 30 Probability: 1.000 AA1: | MAHKRFLKTAFSVLSAVVFFGLVFTSCGG |
| 98349 | 23 AA2: 24 Probability: 0.998 AA1: | MKTSLLKALLLCAVVALFAAGFA |
| 98392 | 29 AA2: 30 Probability: 0.541 AA1: | MTSQIPLWIIPFFPLVGALLLGAITLAQA |
| 98425 | 30 AA2: 31 Probability: 0.988 AA1: | LSIASTSVRRPVLTLVLSAFIVLFGLIGLK |
| 98462 | 27 AA2: 28 Probability: 0.914 AA1: | MKKNIKYVTFALLATILAFIGCGKSNK |
| 98474 | 26 AA2: 27 | MSGRVLRFAPFALACGLALLPIGLNA |
| 98479 | Probability: | GKALKKSPGNSSGFQATEFKSKTKALLGTAYLKSRHSQA |

| | | |
|---|---|---|
| | 0.633 AA1: 39 AA2: 40 Probability: 1.000 AA1: | |
| 98480 | 22 AA2: 23 Probability: 0.999 AA1: | MLYMAVPALVSIIALFCVPAGA |
| 98490 | 26 AA2: 27 Probability: 1.000 AA1: | MIWASNYNRYATATMFSLFFAGKTYA |
| 98495 | 26 AA2: 27 Probability: 1.000 AA1: | MKKILLFGCISMLFSALSCASKGAIA |
| 98509 | 21 AA2: 22 Probability: 1.000 AA1: | MMRRIKKSVVLVLAAVGLVFA |
| 98510 | 27 AA2: 28 Probability: 0.608 AA1: | MKSEFAKLAISATFGLALALMHTAAFA |
| 98511 | 25 AA2: 26 Probability: 1.000 AA1: | VKYGMFIAKRICFIIIFGMVSVGFS |
| 98520 | 24 AA2: 25 Probability: 1.000 AA1: | MRRNSVFVFIFFVAILCIPQYTAA |
| 98554 | 22 AA2: 23 Probability: 0.658 AA1: | MRRNIYIALCALTALCFSNCVA |
| 98565 | 35 AA2: 36 Probability: 0.677 AA1: | SRLFPARYVAALFLLVIKEALVGLAVGFASSFLFT |
| 98574 | 29 AA2: 30 Probability: 1.000 AA1: | EPIPPFIRRPVATTLLTVAIALTGSIAYR |
| 98605 | 23 AA2: 24 Probability: 0.980 AA1: | MRLKSKTLLFLCVIFFAAAPAIM |
| 98614 | 29 AA2: 30 Probability: 0.639 AA1: | MEKTVKVIMIAIVAAVIVCVTVIGLSLAA |
| 98635 | 20 AA2: 21 Probability: 0.725 AA1: | GKDVCFIALGSLMVNQSLKA MILTIRGHLTKDDVKKLKGLLKVVKENKKGP |
| 98637 | 60 AA2: 61 Probability: 0.959 AA1: | VKFIPIIFAASITAAVILFCTIFASPLLG |
| 98638 | 19 AA2: 20 Probability: 0.926 AA1: | MKRYITFLFICIFPVCIYA |
| 98645 | 24 AA2: 25 | AATKILKGIGIGISLVAITRSVVG |
| 98646 | Probability: | MQIQTRAAAAAIDIGTTAIQA |

| | | |
|---|---|---|
| | 0.663 AA1: | |
| | 21 AA2: 22 | |
| | Probability: | |
| | 1.000 AA1: | |
| 98655 | 19 AA2: 20 | MKKSLLACGLLLCVSMALA |
| | Probability: | |
| | 0.993 AA1: | |
| 98657 | 30 AA2: 31 | MQHFKAIRTRTRLTISIILGFVLVSFGCTS |
| | Probability: | |
| | 0.593 AA1: | |
| 98659 | 16 AA2: 17 | MKTFTGIPVSTGVAYA |
| | Probability: | |
| | 1.000 AA1: | |
| 98698 | 21 AA2: 22 | MKKNYTMAAMLFAVSIIGANA |
| | Probability: | |
| | 1.000 AA1: | |
| 98701 | 30 AA2: 31 | VNNFWKTVSGSALAALALVCVFGIAAAVSA |
| | Probability: | |
| | 0.559 AA1: | |
| 98709 | 34 AA2: 35 | MAIFKYEIRQAGVFVAIWAVICAILIFTLFPNFA |
| | Probability: | |
| | 0.555 AA1: | |
| 98718 | 33 AA2: 34 | LAIPNKLAQFTKTLALMGSIGIFVIAALRLVPA |
| | Probability: | |
| | 0.610 AA1: | KTKWSFWKLFVYSLDGIIACSSKPLAIASVLGILLF |
| 98727 | 54 AA2: 55 | LAALLFIIFIIIRRIVWG |
| | Probability: | |
| | 0.532 AA1: | |
| 98764 | 32 AA2: 33 | MAIVPILLRLAGGLCMFLFGMKMMSDGIQQSA |
| | Probability: | |
| | 1.000 AA1: | |
| 98781 | 23 AA2: 24 | MIKKIIHGFFFVLTLSFSLVFSA |
| | Probability: | |
| | 0.613 AA1: | |
| 98792 | 27 AA2: 28 | MKLALIIGIGSFIGGFSRYLLSLLIQA |
| | Probability: | |
| | 0.997 AA1: | |
| 98810 | 37 AA2: 38 | MNYKWIWFIAARYIFRKRKISPAPVLSILGIATGVLA |
| | Probability: | |
| | 1.000 AA1: | |
| 98829 | 29 AA2: 30 | MKTNILGMIKLGLILAVFAAAACVMLAFV |
| | Probability: | |
| | 0.574 AA1: | |
| 98866 | 28 AA2: 29 | MLKRTISSTILKTSSAWPVLFLTGPRQV |
| | Probability: | |
| | 0.995 AA1: | |
| 98868 | 25 AA2: 26 | ANTSVISFLLAVSAAAGFIASPLIG |
| | Probability: | |
| | 1.000 AA1: | |
| 98890 | 19 AA2: 20 | MKKICLILAALVCAVAAQA |
| 98892 | Probability: | MRKSGIIKIIFILFFITSCATISTKTYQ |

| | | |
|---|---|---|
| | 0.986 AA1: 28 AA2: 29 Probability: 0.994 AA1: | |
| 98894 | 18 AA2: 19 Probability: 0.995 AA1: | MKYNMLILIALLAGHCFG |
| 98902 | 27 AA2: 28 Probability: 0.997 AA1: | MRSCFLPVFYFFPIVLLLCLFSTEAGA |
| 98931 | 16 AA2: 17 Probability: 0.995 AA1: | KLIRGCLALAIVVSMA |
| 98938 | 30 AA2: 31 Probability: 0.987 AA1: | MNSLPKRRFLFAVSALFAAALYILIRYAMR |
| 98952 | 20 AA2: 21 Probability: 0.740 AA1: | MKRIVAVIFIAVFSALSITG |
| 98964 | 33 AA2: 34 Probability: 0.834 AA1: | ASTLTTALVFVPIAFLPGPLGSLFGDMAVSLVT MGFSRGFSRKPVFIEALRYSIPVLLGYVTLGFAFG |
| 98967 | 69 AA2: 70 Probability: 0.995 AA1: | VAAGTGYPWWLALFMSVWMFAGAGQFFAIGLFA |
| 98968 | 19 AA2: 20 Probability: 1.000 AA1: | MKINSVLLAGAGAVGLTVA |
| 98995 | 22 AA2: 23 Probability: 0.869 AA1: | VNKKPETGLSSGALLSAAGAQA |
| 98996 | 23 AA2: 24 Probability: 0.938 AA1: | MCAIGGHPLLAIPAGLIAGAAAG |
| 99009 | 22 AA2: 23 Probability: 0.703 AA1: | EKSGRAALAAHAILGASAPENA |
| 99015 | 19 AA2: 20 Probability: 0.995 AA1: | MAAPQKPLALVFMGTPLFA |
| 99027 | 32 AA2: 33 Probability: 0.516 AA1: | MILSNLKNCSIVLASVLGLAFVGCGGPSGTDA |
| 99040 | 22 AA2: 23 Probability: 0.993 AA1: | MFRRKRHIHFVGIGGIGMCGLA |
| 99070 | 24 AA2: 25 Probability: 0.998 AA1: | MLSGMKPYVALIGIALFLALFAVA |
| 99080 | 22 AA2: 23 | MMKQIFTLLAAAAVFASCQRKA |
| 99114 | Probability: | MRRVVITGSSVISALG |

| | | |
|---|---|---|
| | 0.508 AA1: 16 AA2: 17 Probability: 1.000 AA1: | |
| 99122 | 24 AA2: 25 Probability: 0.992 AA1: | MKKIILLTVALSLLVFLSAGSSES |
| 99159 | 27 AA2: 28 Probability: 0.971 AA1: | MNARSLLIIPALILILTAGGLSGRAPA |
| 99160 | 20 AA2: 21 Probability: 0.635 AA1: | ANRFCVLLGSALLSVWGIWA<br>MKNITRDGRNEFRRDRPGKNSGPAFSFL |
| 99165 | 45 AA2: 46 Probability: 0.999 AA1: | HKIVPALLIAASLIAAT |
| 99182 | 25 AA2: 26 Probability: 0.678 AA1: | MAIAGASRFPVFILSALLWAAGASA |
| 99202 | 18 AA2: 19 Probability: 0.585 AA1: | MPAILNLIWLLPILAVFA |
| 99217 | 24 AA2: 25 Probability: 0.862 AA1: | MQKLKVAVCGVGSLGQHHARVVAA |
| 99230 | 20 AA2: 21 Probability: 1.000 AA1: | LQSRLGLLPSVNASASFSSS |
| 99267 | 20 AA2: 21 Probability: 0.999 AA1: | MKKIITTVIAVLIYAGAVFA |
| 99271 | 37 AA2: 38 Probability: 0.642 AA1: | MNKEQLAMSNGQRAMSVFARLIAIFLLLCGICLNAVA |
| 99274 | 17 AA2: 18 Probability: 0.761 AA1: | FGIGSLLIACIGSCSNA |
| 99284 | 26 AA2: 27 Probability: 0.683 AA1: | MRFLNGVLLVLLLSSCIGRVVTPIFE<br>LLAAGRFLTRRYTVVFVPHSEKRVYNFHINVLSI |
| 99328 | 46 AA2: 47 Probability: 0.992 AA1: | LCFLLVAGAFMG |
| 99329 | 33 AA2: 34 Probability: 0.525 AA1: | MKLSVRIPLLIGVVVLITSAVLGILAFQLSSKA<br>MKIKNSGVSWIPWLAPALVFYVFFMAWPLLDSLR |
| 99335 | 49 AA2: 50 Probability: 0.726 AA1: | LSLFTGSAGMGRTFA |
| 99365 | 24 AA2: 25 | MATHFGRIILPAQLFFFAGSFLIA |
| 99370 | Probability: | MNLFLPVFACLTLAGGPACFG |

| | | |
|---|---|---|
| | 0.992 AA1: 21 AA2: 22 Probability: 0.920 AA1: | |
| 99373 | 22 AA2: 23 Probability: 0.994 AA1: | MKISFGKAMCTGFLCAAVLLTA |
| 99374 | 21 AA2: 22 Probability: 1.000 AA1: | MKRFNLILLIVLFGFLFSCAG |
| 99402 | 19 AA2: 20 Probability: 0.587 AA1: | MKKILLFTLCGLLLNFAWA |
| 99408 | 36 AA2: 37 Probability: 0.574 AA1: | MTWYGGILTAAVAAFPLMYRTARGAFEGFDNTLAHA |
| 99409 | 22 AA2: 23 Probability: 0.666 AA1: | MLAALAQTVSSAQSLLAGSANG MGPDMTKKIKVVKRFLKQAKFTNLV |
| 99428 | 38 AA2: 39 Probability: 0.985 AA1: | LYAALVLISYQFM |
| 99442 | 27 AA2: 28 Probability: 0.991 AA1: | MKSLNYMKIVIFVVSISQFLFSLNLIA |
| 99451 | 23 AA2: 24 Probability: 0.965 AA1: | MRHKFALPLLLILFSSCRTISHS |
| 99460 | 23 AA2: 24 Probability: 0.511 AA1: | VWLSGSTLALSLLMVAGLVLLLA |
| 99462 | 19 AA2: 20 Probability: 0.867 AA1: | MFSTSSLVTAFPAAQTGMG |
| 99463 | 28 AA2: 29 Probability: 1.000 AA1: | MTLFTGIRNQDYPLISGSTLIITVMVLA |
| 99467 | 27 AA2: 28 Probability: 0.910 AA1: | MKIKMKDKLFTFLLAAAFLLMPVSLFA |
| 99503 | 33 AA2: 34 Probability: 0.994 AA1: | MSMVKTVVSRPTTFLIIFLLLIGLGFFAMFNLS |
| 99509 | 17 AA2: 18 Probability: 0.573 AA1: | VRRRLYLLLICSHCVFA |
| 99530 | 31 AA2: 32 Probability: 0.990 AA1: | VTHSIPGTYAVSLAVPSAGNIFVSADFKAFA |
| 99544 | 34 AA2: 35 | MNLRDTIKNFKWNEHSTGLGFIMLFVLGAIVSWA |
| 99567 | Probability: | MDVQKRLYYGIKRIEMKQITLLALIASLCALNAPA |

| | | |
|---|---|---|
| | 0.996 AA1: 35 AA2: 36 Probability: | |
| | 0.842 AA1: | |
| 99603 | 31 AA2: 32 | ILLGLAVMTTGLGLLGLPAGIGAFLAGMLFS |
| | Probability: 0.998 AA1: | |
| 99627 | 31 AA2: 32 | VNRKNVPGCRPFLRLLIVILLVFFGASMLPA |
| | Probability: 1.000 AA1: | |
| 99633 | 28 AA2: 29 | MSFFNQRGRTLAFAAATVIGAVCFSVNA |
| | Probability: 0.996 AA1: | |
| 99660 | 27 AA2: 28 | MGFNGFSINMMTLLALASALGTLIANA |
| | Probability: 0.979 AA1: | |
| 99710 | 25 AA2: 26 | LPYGFLIGVLGFGFFLICAFSVFFA |
| | Probability: 1.000 AA1: | |
| 99732 | 19 AA2: 20 | MSLKALFLAVALCLGASFA |
| | Probability: 0.549 AA1: | |
| 99765 | 20 AA2: 21 | DGSTAAVTGYTTSGVSSTTA |
| | Probability: 0.616 AA1: | |
| 99775 | 28 AA2: 29 | MGRTPGIICVCSFLAAVIVTLTNLNLSS |
| | Probability: 0.995 AA1: | |
| 99790 | 38 AA2: 39 | MVNILKKASPMKRRASLTDSNLLFLIAICVFVILYALA |
| | Probability: 0.796 AA1: | |
| 99820 | 25 AA2: 26 | MLAVLAVIILVISIVLFMQGHSLLV |
| | Probability: 0.999 AA1: | |
| 99823 | 23 AA2: 24 | MRKHILCLMATIFALAFTPIACG |
| | Probability: 0.936 AA1: | |
| 99845 | 35 AA2: 36 | MEASSAVLPITAIVLIASMVLPMPSGTILMFLAGA |
| | Probability: 0.983 AA1: | |
| 99877 | 20 AA2: 21 | MKRISVILIITCVINVSVFG |
| | Probability: 0.761 AA1: | |
| 99884 | 22 AA2: 23 | MIVKKAFIISLVCVLFIGCVKN |
| | Probability: 0.996 AA1: | |
| 99919 | 22 AA2: 23 | MKMKAILLTLIISTALIGTCAT |
| | Probability: 1.000 AA1: | |
| 99922 | 25 AA2: 26 | MITAFRKVVPILAALLILLPTASQA |
| 99945 | Probability: | LFITLLGIAASTIGMFEETLPFIPIFIGFAIA |

| | | |
|---|---|---|
| | 0.802 AA1: 32 AA2: 33 Probability: 0.951 AA1: | |
| 99955 | 31 AA2: 32 Probability: 0.996 AA1: | MKGIKPVLLAISTALVLKLFFFDFIIAQGHS |
| 99962 | 26 AA2: 27 Probability: 1.000 AA1: | MITRYMQIVFVALAFALMIIASYLFV |
| 99968 | 23 AA2: 24 Probability: 1.000 AA1: | MKIKLRLSLAMIALMAVVVTVIA |
| 99973 | 28 AA2: 29 Probability: 0.545 AA1: | MRGKNKNKWVFFAVLLLISLILPTKIAA |
| 99979 | 17 AA2: 18 Probability: 0.995 AA1: | MISLVILAALYVVMTFA |
| 99997 | 29 AA2: 30 Probability: 0.731 AA1: | GIRPFHLQMLFVALAFIAMVVSSTFQVNA IAPLVFYESYTSDKKTLWNMRRPILGAAIMLV |
| 100008 | 53 AA2: 54 Probability: 0.716 AA1: | LITVIAAGYITNFFIPAIPLA |
| 100024 | 19 AA2: 20 Probability: 1.000 AA1: | VADLMLVLGTSLAVFPAAG |
| 100044 | 29 AA2: 30 Probability: 0.879 AA1: | MNKTVRTVVWSVMMAVIAVLTVGAAGAWA |
| 100045 | 24 AA2: 25 Probability: 0.847 AA1: | MISKIYIGFAVPLAATSLTFISMM |
| 100049 | 23 AA2: 24 Probability: 0.983 AA1: | LNTERKVLVTGASGGIGRAIALA |
| 100072 | 22 AA2: 23 Probability: 0.993 AA1: | MKINNINKFIIAVLCGAVFIAA |
| 100086 | 37 AA2: 38 Probability: 0.625 AA1: | MKLFHSLRFRFAFIFSVFIIVLSVSLSLLGIRQTTKA |
| 100093 | 24 AA2: 25 Probability: 0.588 AA1: | MPIMKIFTGRCPLVLIICSLVLSS |
| 100103 | 19 AA2: 20 Probability: 0.800 AA1: | MILCSLLAVIGNVAANRMS GTDVKSKAQVTKTIFRLWRYLSRYRLRIVIVISLTVLS |
| 100104 | 52 AA2: 53 | NFLALLGPMLSGRA |
| 100105 | Probability: | MSNEQWKRKKITMLRLTAFCLLLAVISPMAWA |

| | | |
|---|---|---|
| | 1.000 AA1: | |
| | 32 AA2: 33 | |
| | Probability: | |
| | 0.998 AA1: | |
| 100106 | 23 AA2: 24 | MKKYLLIPLLAALFIACGQKEQA |
| | Probability: | |
| | 1.000 AA1: | |
| 100110 | 21 AA2: 22 | MKSKAFVLVSVTFFLALPLFA |
| | Probability: | |
| | 0.863 AA1: | |
| 100124 | 27 AA2: 28 | MLLRALCAISLYILACSSKDSGLVVYG |
| | Probability: | |
| | 0.994 AA1: | |
| 100136 | 20 AA2: 21 | MKRVLIITVMLCFMAAAIFG |
| | Probability: | |
| | 0.806 AA1: | |
| 100145 | 33 AA2: 34 | MTRATSVGFLLLSPICTSLTGMFIKFIPWSPWA |
| | Probability: | |
| | 0.973 AA1: | |
| 100156 | 19 AA2: 20 | VPIILIPTIILLAFALANA |
| | Probability: | |
| | 0.993 AA1: | |
| 100205 | 21 AA2: 22 | MRKKKTLFLCLLLLIINKIHV |
| | Probability: | |
| | 0.997 AA1: | |
| 100210 | 28 AA2: 29 | MKLNNNEKITKKLSMAFGLAVTFGVVMS |
| | Probability: | |
| | 0.773 AA1: | |
| 100211 | 20 AA2: 21 | MHDILVIGAGIAGLTVALKA |
| | Probability: | |
| | 0.958 AA1: | |
| 100215 | 24 AA2: 25 | MSIKGKAAVVTGSATGIGQGVALA |
| | Probability: | |
| | 0.903 AA1: | |
| 100218 | 17 AA2: 18 | FIAILIFGGIFVFPALS |
| | Probability: | |
| | 0.503 AA1: | |
| 100234 | 25 AA2: 26 | MARLEELTVGVSVAGLAGSASVSIV |
| | Probability: | |
| | 0.978 AA1: | |
| 100250 | 21 AA2: 22 | LKAAIGLSAGTLTIVSTPTAA |
| | Probability: | |
| | 0.944 AA1: | |
| 100257 | 34 AA2: 35 | MLQFRMALNCGIVGLPNVGKSTIFSALSSAPAEA |
| | Probability: | |
| | 0.999 AA1: | METLLTSFVSLFNSWSTLGVMVFFLVACFL |
| 100261 | 40 AA2: 41 | LGMPIFSALA |
| | Probability: | |
| | 0.985 AA1: | |
| 100281 | 29 AA2: 30 | MNNIKIGVKLIFSFLLVSVLATIMGIYMS |
| 100294 | Probability: | MILTLFIGLISVSPVPSLTFLVVLGLGGALA |

| | | |
|---|---|---|
| | 0.893 AA1: | |
| | 31 AA2: 32 | |
| | Probability: | |
| | 0.628 AA1: | |
| 100296 | 27 AA2: 28 | MKTQNLLKSVLDAAIFGGLLLMASCDS |
| | Probability: | |
| | 0.893 AA1: | |
| 100307 | 23 AA2: 24 | TSLKLPLSTTYVTFMVAMGTSLS |
| | Probability: | |
| | 0.994 AA1: | |
| 100319 | 21 AA2: 22 | LKYSNPATLAIIACPGGEVFA |
| | Probability: | |
| | 0.560 AA1: | MGKEKTAKIQADKISRPIGAKLITIITILLLFS |
| 100324 | 46 AA2: 47 | LGAITVLVSVLVS |
| | Probability: | |
| | 0.924 AA1: | |
| 100331 | 18 AA2: 19 | MNVMFTALILFISAALYG |
| | Probability: | |
| | 0.823 AA1: | |
| 100341 | 24 AA2: 25 | SERPYLSSAVPIVCLLLIAACVFG |
| | Probability: | |
| | 0.641 AA1: | |
| 100348 | 42 AA2: 43 | MLFYDMLIDGIKTKKAGKIAAAVLLCLVPIITSLPLIFVANL |
| | Probability: | |
| | 0.982 AA1: | |
| 100362 | 23 AA2: 24 | LSLRKGLIIFILLAFGANVLILA |
| | Probability: | |
| | 0.576 AA1: | |
| 100381 | 23 AA2: 24 | MTDKRLGYIVMFLILLLVGGVYG |
| | Probability: | |
| | 0.523 AA1: | |
| 100387 | 19 AA2: 20 | DIIILATASADYFGTPSTA |
| | Probability: | |
| | 1.000 AA1: | MNNNGFSDRNKPSRRRIPFSWIWFGSTLLL |
| 100392 | 44 AA2: 45 | CAVFAFASVPSVLA |
| | Probability: | |
| | 1.000 AA1: | |
| 100412 | 39 AA2: 40 | MKRKINLGYVLIILFVGLFAFLCLYPFLLVISGSFTTQA |
| | Probability: | |
| | 0.923 AA1: | |
| 100432 | 20 AA2: 21 | MNKFIFVSGGVCSSLGKGVA |
| | Probability: | |
| | 0.999 AA1: | |
| 100439 | 26 AA2: 27 | MRTQLKKIALTATFGLAMALNFSCSS |
| | Probability: | |
| | 0.998 AA1: | |
| 100440 | 20 AA2: 21 | MGVGRGIRALALLAAFAVVA |
| | Probability: | |
| | 0.658 AA1: | |
| 100443 | 21 AA2: 22 | MERAFASFTALALSLNIPALK |
| 100446 | Probability: | MTLDRYLLKHFFPVFLAGLILFAMMVLMI |

| | | |
|---|---|---|
| | 0.901 AA1: 46 AA2: 47 Probability: | DLFLNLTRYLSNGAAFS |
| | 0.994 AA1: | MVILLLFLRNIKATVIIGLSIPISMLLTLMSMSIF |
| 100451 | 57 AA2: 58 Probability: | GFTLNLMTMTGLIMGMGMTLDA |
| | 0.997 AA1: | |
| 100518 | 20 AA2: 21 Probability: | MILKKIAVFFLLFAVLCCMA |
| | 0.570 AA1: | |
| 100519 | 38 AA2: 39 Probability: | MKSGVFANEDKHGFSPQRKVFFFSLLFFFVILAGGSLA |
| | 1.000 AA1: | |
| 100526 | 24 AA2: 25 Probability: | MKKNQLFISLLAAAFLFLAAPVFA |
| | 0.931 AA1: | |
| 100530 | 36 AA2: 37 Probability: | LSCANPYREIFVPLAALAGAGSGLTSSIGLTSGAGA |
| | 0.778 AA1: | |
| 100552 | 28 AA2: 29 Probability: | MTLVLWGAGLVMGIAAIGSGFGIGIAGQ |
| | 1.000 AA1: | |
| 100563 | 18 AA2: 19 Probability: | VRVIFLLLLIFASVNVFA |
| | 0.966 AA1: | |
| 100572 | 26 AA2: 27 Probability: | MNLGLIGAGLVMGISAVGSGMGIAIA |
| | 1.000 AA1: | |
| 100575 | 19 AA2: 20 Probability: | MKRFRFLAAAFLVVTLAWA |
| | 0.977 AA1: | |
| 100609 | 30 AA2: 31 Probability: | MFLRKFVVKPLHFFGVVGLLLCLAATGVFG |
| | 0.518 AA1: | |
| 100647 | 20 AA2: 21 Probability: | ANGSAGGIVPGACVTGLAGT |
| | 0.832 AA1: | |
| 100674 | 21 AA2: 22 Probability: | GKFRHADIALSTAVHWMPSNA |
| | 0.999 AA1: | |
| 100677 | 20 AA2: 21 Probability: | MNKKIFIACVCLGFVGSIAA |
| | 0.729 AA1: | |
| 100692 | 20 AA2: 21 Probability: | VFLTAGMGGGTGTGATPVIA |
| | 0.997 AA1: | |
| 100696 | 21 AA2: 22 Probability: | MKKTGSIIFIVLILVSWNVFS |
| | 0.987 AA1: | |
| 100724 | 25 AA2: 26 | VYKPDIPLFVLALVFVFLGVLLVYG |
| 100725 | Probability: | MIPKKWNVERKVIYMTILLLTVFFLLPFVLFG |

| | | |
|---|---|---|
| | 0.981 AA1: | |
| | 32 AA2: 33 | |
| | Probability: | |
| | 1.000 AA1: | |
| 100728 | 22 AA2: 23 | MKTIMKICTVCLLLASSTTLFA |
| | Probability: | |
| | 0.629 AA1: | |
| 100733 | 22 AA2: 23 | PTGATGPAGATGPTGATGPAGA |
| | Probability: | |
| | 0.808 AA1: | |
| 100736 | 26 AA2: 27 | NPNQTNYTSGTAVTVTATAVSGYVFT |
| | Probability: | |
| | 0.995 AA1: | |
| 100780 | 21 AA2: 22 | DLIFQTVFCATAATIVSGAMA |
| | Probability: | |
| | 1.000 AA1: | |
| 100785 | 21 AA2: 22 | MKKNSIFLLIALFAFSFPGFA |
| | Probability: | |
| | 0.998 AA1: | |
| 100813 | 16 AA2: 17 | MNKIATLLLLCTTLFA |
| | Probability: | |
| | 0.988 AA1: | |
| 100825 | 25 AA2: 26 | MGKKNITYLAVLAALLTPAIVSCNS |
| | Probability: | |
| | 0.995 AA1: | |
| 100832 | 17 AA2: 18 | MNKIITLLALLNIAVFA |
| | Probability: | |
| | 0.804 AA1: | |
| 100837 | 24 AA2: 25 | VFVDDFRPMFFCLRSTPASATAAA |
| | Probability: | |
| | 0.587 AA1: | MLDKLFPKKTAPVAPTSQPALSLGGLVS |
| 100838 | 38 AA2: 39 | GIKSAASGLA |
| | Probability: | |
| | 1.000 AA1: | |
| 100846 | 23 AA2: 24 | MKNLFKLIGIIALTALIGFSFLA |
| | Probability: | |
| | 0.868 AA1: | |
| 100849 | 21 AA2: 22 | MKKAVVLLSGGLDSATCGVVA |
| | Probability: | |
| | 0.833 AA1: | |
| 100862 | 24 AA2: 25 | GKLNQTAGSVTAVTITPISGKSTG |
| | Probability: | |
| | 0.996 AA1: | |
| 100870 | 28 AA2: 29 | NAVSSTQVKARLGIFAAMVLCFAYGAWA |
| | Probability: | |
| | 0.998 AA1: | |
| 100876 | 19 AA2: 20 | MKKYITVFLCVALAALTVA |
| | Probability: | |
| | 1.000 AA1: | |
| 100884 | 27 AA2: 28 | MKKKVLMSLVLLAFVTASAIFAPRVFA |
| 100891 | Probability: | MPLWASFSFSICANVGLTMTAAFNVFA |

| | | |
|---|---|---|
| | 0.997 AA1: 27 AA2: 28 Probability: 0.915 AA1: | |
| 100894 | 25 AA2: 26 | MKTKPWIIPVISFIVYLLAVPALFG |
| | Probability: 0.886 AA1: | MEDRIRSNNPRDFKKNRHLYLMMLPFML |
| 100898 | 40 AA2: 41 | LFFVFTVLPVLT |
| | Probability: 1.000 AA1: | |
| 100947 | 19 AA2: 20 | MLTLLLALASILSSRMLAA |
| | Probability: 1.000 AA1: | |
| 100951 | 20 AA2: 21 | MRRLVLITAAVAAAIAGAFA |
| | Probability: 0.999 AA1: | |
| 100968 | 20 AA2: 21 | VNGFTKIALVAAFGFALCFA |
| | Probability: 0.987 AA1: | |
| 100992 | 31 AA2: 32 | MNRFTFKMFGLGRTLLLTAVLIVGAVCVSYG |
| | Probability: 1.000 AA1: | |
| 101009 | 24 AA2: 25 | LKRFSLLIALFLSLLLWAAGFVYA |
| | Probability: 0.869 AA1: | |
| 101028 | 20 AA2: 21 | LIAVAVTVVVAVIDTVPASA |
| | Probability: 0.999 AA1: | |
| 101040 | 16 AA2: 17 | VFAVCCMLSAASAVSA |
| | Probability: 0.976 AA1: | |
| 101045 | 26 AA2: 27 | MKERIKKFLFVCLILALFVSCPPDSE |
| | Probability: 0.996 AA1: | |
| 101059 | 34 AA2: 35 | MKRAMPENKGRRKPYHLSFVIFYLLFFICSAAWA |
| | Probability: 0.798 AA1: | |
| 101062 | 31 AA2: 32 | MRKFYNYPWLIVAAIAIVTVFFGFQLPRAQL |
| | Probability: 1.000 AA1: | |
| 101070 | 35 AA2: 36 | MRMIDCRPRRPAFTAAALVVILGCAAAAFPPAARA |
| | Probability: 0.568 AA1: | |
| 101077 | 43 AA2: 44 | MRIKKVYVNSKIKTASDGIYALGLLLPYGLLFSVFIAIPIAVA |
| | Probability: 0.639 AA1: | |
| 101087 | 28 AA2: 29 | YVFGAILVFLFLAVCRLFAPFFSVLLWS |
| | Probability: 1.000 AA1: | |
| 101090 | 19 AA2: 20 | MKKIFVVIFLLISIMFSFA |
| 101093 | Probability: | MMTFLNKALSFLTRLGSRADLTVPVFVLL |

| | | |
|---|---|---|
| | 0.868 AA1: 42 AA2: 43 Probability: 0.995 AA1: | AIVVMILPMPAWG |
| 101109 | 23 AA2: 24 Probability: 0.670 AA1: | MKFIGKALFPIILFTGMAGIGHA |
| 101119 | 34 AA2: 35 Probability: 0.822 AA1: | VSWNNPTGGTITATAVIAPSSTPAPYTNGTASLA |
| 101159 | 23 AA2: 24 Probability: 0.981 AA1: | LPGKKRILILGAGVLQGPALRLA |
| 101164 | 22 AA2: 23 Probability: 0.999 AA1: | MIRLLCIIIAGVTLFMLPSLFT |
| 101170 | 22 AA2: 23 Probability: 1.000 AA1: | MFHAALVLIIMVVILAIGSVLA |
| 101171 | 28 AA2: 29 Probability: 0.549 AA1: | MKNKYTGFAVLLLVLVSLAIFVSSLSFM |
| 101178 | 20 AA2: 21 Probability: 0.822 AA1: | MKALFIGGTGTISSAISRLA |
| 101188 | 24 AA2: 25 Probability: 0.984 AA1: | LKVLKLFVFIAFFALAFIHEPCRA |
| 101212 | 29 AA2: 30 Probability: 0.640 AA1: | MFLETSFYHLKILLPIIALCCGSIFALVA |
| 101252 | 26 AA2: 27 Probability: 0.990 AA1: | DLPRLPYYQSESKLAAIFSILGIALA |
| 101283 | 24 AA2: 25 Probability: 0.997 AA1: | MSNKSVFKNAAAVLLAALFLIISG<br>NYEKRFHLNNKPRGLVGAFLAGGAFGAGW |
| 101295 | 38 AA2: 39 Probability: 0.804 AA1: | TPCIGPALA |
| 101297 | 32 AA2: 33 Probability: 0.923 AA1: | MLNAANINFIAFAHFRALDATAGAIFSLFVIA |
| 101333 | 20 AA2: 21 Probability: 0.999 AA1: | AAVAAVAVVIAVPVSVRAEA |
| 101335 | 20 AA2: 21 Probability: 0.993 AA1: | MKIKIKLSIVVIAIMAVVVA |
| 101340 | 29 AA2: 30 | MNKIKRPERNRFVFAAVFAALIAVSNFFQ |
| 101346 | Probability: | MKKPFVVLVFSIAMFSCVK |

| | | |
|---|---|---|
| | 0.773 AA1: 19 AA2: 20 Probability: | |
| | 0.998 AA1: | |
| 101377 | 32 AA2: 33 Probability: | VKKTFIPIISAFLALAPCTISCGGGGNAPAHA |
| | 0.999 AA1: | |
| 101380 | 23 AA2: 24 Probability: | TGILFLLALFLSPLFLLIPSVAT |
| | 0.830 AA1: | |
| 101385 | 18 AA2: 19 Probability: | SIKTAFLICAYLTSVTFQ |
| | 0.999 AA1: | |
| 101387 | 22 AA2: 23 Probability: | GRKACALGVPCLTSIDTANALA |
| | 0.988 AA1: | |
| 101401 | 23 AA2: 24 Probability: | MKVSKKALTALAVITVITLAGCA |
| | 1.000 AA1: | |
| 101417 | 20 AA2: 21 Probability: | MLRKKISLTAALVCCVIAFA |
| | 0.993 AA1: | |
| 101433 | 29 AA2: 30 Probability: | MRRRQVKKTVLLFTFLFTSIFCFIQCSSG |
| | 0.694 AA1: | |
| 101437 | 27 AA2: 28 Probability: | MTYFIHAIMIFLGLLAVVPIYFLLLNA |
| | 1.000 AA1: | MRRGRYIRGQRNKFNPKTRLIGMNTKQLK |
| 101452 | 51 AA2: 52 Probability: | KSIWPAAVTAVTLTLAPAPVLS |
| | 0.860 AA1: | |
| 101510 | 22 AA2: 23 Probability: | MKLLLHCCCAPCSVSCIGSLRA |
| | 0.524 AA1: | MTLSGIWELKKIAMRNLARHKVKTILTSAAI |
| 101514 | 41 AA2: 42 Probability: | MVSVTTFIFM |
| | 0.986 AA1: | |
| 101520 | 29 AA2: 30 Probability: | MNNIKIGVKLIASFLLVSILATVMGIYMS |
| | 0.957 AA1: | |
| 101592 | 21 AA2: 22 Probability: | VALLDASIIGAMAVSVSTAYA |
| | 0.915 AA1: | |
| 101611 | 23 AA2: 24 Probability: | MSKKSKFLAFLFCLFFGVFGVHK |
| | 0.913 AA1: | |
| 101613 | 34 AA2: 35 Probability: | LLSCPRGSVQKTFHFLSSVKFTVALLVLLMVGLA |
| | 0.513 AA1: | |
| 101639 | 21 AA2: 22 | MDTITVCMGSSCFGKGNSANA |
| 101640 | Probability: | TALPAKKRAALLRALGRALA |

| | | |
|---|---|---|
| | 0.760 AA1: 20 AA2: 21 Probability: 0.950 AA1: | |
| 101667 | 35 AA2: 36 Probability: 0.995 AA1: | MGGIIDKTKTKFGKARPWLLWMAGPFALASFLLFA FFGGRKRNENKLWAKIFKNKAVSSLFTLLPAYLLAIM |
| 101670 | 54 AA2: 55 Probability: 0.972 AA1: | GYQSIWALFGAANQLLA |
| 101681 | 27 AA2: 28 Probability: 0.993 AA1: | MNGSSAGTFAVLVSGDPGFFSAASGLA |
| 101684 | 28 AA2: 29 Probability: 0.920 AA1: | MSLLYNVFMWIIFALLSAFFAALTSILA |
| 101721 | 22 AA2: 23 Probability: 0.873 AA1: | MRNKLFIAGIIIILIFGIIGMG |
| 101724 | 31 AA2: 32 Probability: 0.784 AA1: | TLSTLSLSVSAEISASVALSISSSLTLSDNA |
| 101733 | 31 AA2: 32 Probability: 0.899 AA1: | MSDRMHMFVPLIAYIALTLCVALWSRRSASS |
| 101743 | 34 AA2: 35 Probability: 0.801 AA1: | MGILDSIKKTAADSLKKAAATTVNTAAATSKAAA |
| 101751 | 19 AA2: 20 Probability: 1.000 AA1: | LRFWEIFLIGVGLSMDAFA |
| 101758 | 30 AA2: 31 Probability: 0.911 AA1: | MNKIRFYRLTVLLCLFALAPVGIMAQNTAA MKGAMTKSREPRVENKTGIKTLIFVICYLLFV |
| 101759 | 40 AA2: 41 Probability: 1.000 AA1: | ICGSAVWA |
| 101769 | 23 AA2: 24 Probability: 0.789 AA1: | MSNINRFFSLALMAFFIVSAAAA |
| 101770 | 32 AA2: 33 Probability: 0.991 AA1: | VNIFTHFSIDNKWKLLIITGFLGGYTTFSTYS |
| 101789 | 23 AA2: 24 Probability: 0.974 AA1: | LKSKFSNLFWGIALLLVAVFVLT |
| 101800 | 17 AA2: 18 Probability: 0.994 AA1: | LAVILPAVLFFAISAFS |
| 101804 | 18 AA2: 19 | MWVLVAAALVFFMQAGFA |
| 101813 | Probability: | FILAGLSLLTLAAPAFA |

| | | |
|---|---|---|
| | 1.000 AA1: 17 AA2: 18 Probability: 0.981 AA1: | |
| 101814 | 32 AA2: 33 | VKQDSFYNSFFWRLALILAVLWLFITTAPFVS |
| | Probability: 0.827 AA1: | |
| 101821 | 27 AA2: 28 | DAAREIRQALTATTTAGAGGASSGLNS |
| | Probability: 1.000 AA1: | |
| 101829 | 27 AA2: 28 | MKKNNPAKRLVLLAIVAFAAIGLAITA |
| | Probability: 0.797 AA1: | |
| 101846 | 35 AA2: 36 | MKKKTRKTGKVSAVKLLIFSFIIIIIVLGILLSYR |
| | Probability: 0.971 AA1: | |
| 101849 | 23 AA2: 24 | TLAPITLGLGALIAISVFPQKQA |
| | Probability: 0.875 AA1: | |
| 101859 | 31 AA2: 32 | MSPIFIGVIALVLVAAIFLIIKNGASGLSFG |
| | Probability: 0.997 AA1: | |
| 101860 | 17 AA2: 18 | MKKTTIMLLLFCMVCFA |
| | Probability: 1.000 AA1: | |
| 101869 | 26 AA2: 27 | MLLIRKYTLFLICCLISLIAATTVFA |
| | Probability: 0.773 AA1: | |
| 101882 | 33 AA2: 34 | VLIEAFNLQIKNILLGINIVLIIIFMSCGKTNA |
| | Probability: 0.994 AA1: | |
| 101889 | 21 AA2: 22 | MCYPIVAVAAMAVMAAGSAYM |
| | Probability: 0.629 AA1: | |
| 101902 | 23 AA2: 24 | ANIVKIILLGFFTAIFGDLGDLA |
| | Probability: 0.999 AA1: | |
| 101904 | 20 AA2: 21 | MRRILFTLFVGFLFVVQLAA |
| | Probability: 0.996 AA1: | |
| 101914 | 25 AA2: 26 | MRFFDTRLGRCATLLALGGGVALYA |
| | Probability: 0.935 AA1: | |
| 101915 | 27 AA2: 28 | MRNCGLFLLGTLALLVSCHPSGQKIHA |
| | Probability: 0.999 AA1: | |
| 101918 | 18 AA2: 19 | KQNPLLAVLFAISAFSLA |
| | Probability: 0.999 AA1: | |
| 101950 | 31 AA2: 32 | MRQRGVFLAVIVAIVVGIVMLFRVSATPVLA |
| 101952 | Probability: | MKKKKTKKKVLSVALATCALAAA |

| | | |
|---|---|---|
| | 1.000 AA1: 23 AA2: 24 Probability: 0.937 AA1: | |
| 101956 | 27 AA2: 28 | MIKRSQTTVCIAFWLILFGADRVSATG |
| | Probability: 0.508 AA1: | |
| 102000 | 33 AA2: 34 | MQNLKVAMKLFVGFGITVVLMIVVGVVSVVSLK |
| | Probability: 0.624 AA1: | |
| 102004 | 18 AA2: 19 | VYAAGICLIANTAPVAFG |
| | Probability: 1.000 AA1: | |
| 102034 | 23 AA2: 24 | MLFKRTFLVLAGLLIASVSVLWA |
| | Probability: 1.000 AA1: | |
| 102038 | 23 AA2: 24 | MKKRTGICMLLACCLLAACIATK |
| | Probability: 0.733 AA1: | |
| 102040 | 26 AA2: 27 | MRLYSRGQLVFFSIVSCFLTAIIVTG |
| | Probability: 0.863 AA1: | |
| 102087 | 19 AA2: 20 | ITFNATGLGTASGGFSLSG |
| | Probability: 1.000 AA1: | |
| 102104 | 33 AA2: 34 | MKIKCSNLSPICFAIAGVILFACSLGSATPSSS |
| | Probability: 0.999 AA1: | |
| 102127 | 23 AA2: 24 | MKKNRVVVLTAICVLLAACSSDS |
| | Probability: 0.599 AA1: | MDLQVLIAIGAGIAVFTGVGAGIGIGIATSGFLQ |
| 102152 | 57 AA2: 58 | AISRQPEASGKMTPMFFVGVALA |
| | Probability: 1.000 AA1: | |
| 102181 | 28 AA2: 29 | KLISAALLCFIAWFMCRVSLAMPGNSLA |
| | Probability: 0.970 AA1: | |
| 102200 | 21 AA2: 22 | MASKGKVAVALSGGVDSSVVA |
| | Probability: 0.997 AA1: | |
| 102221 | 24 AA2: 25 | MPLGVIVVASVAGLFLALPVIAVA |
| | Probability: 0.732 AA1: | |
| 102225 | 20 AA2: 21 | EVIKHAVCPVSLSAPVIVTG |
| | Probability: 0.994 AA1: | |
| 102252 | 24 AA2: 25 | MNEKTNIRKSVVIGGLICVNSLFA |
| | Probability: 1.000 AA1: | |
| 102276 | 23 AA2: 24 | MLLKYFKFVSFCIIAAFSQIVFA |

| | | |
|---|---|---|
| | Probability: 0.999 AA1: | |
| 102278 | 21 AA2: 22 | MKNIIFLFAGIVFSAALPAQT |
| | Probability: 0.995 AA1: | MKLLKNLLKSPMFVIGASLFLLTLLIALVGPF |
| 102305 | 42 AA2: 43 | FYKITMARVG |
| | Probability: 0.917 AA1: | |
| 102328 | 35 AA2: 36 | VGTPTKAETKGIAIGSAVIAAVSLFSSFIAVIAVG |
| | Probability: 1.000 AA1: | |
| 102337 | 21 AA2: 22 | VASGKRLVASGLILLTLQALA |
| | Probability: 0.999 AA1: | |
| 102346 | 35 AA2: 36 | MASIMTIFKKAKLRTKIISVVIAICLFALLGMSIA |
| | Probability: 0.833 AA1: | |
| 102371 | 18 AA2: 19 | RFSKSLSLGGIVTHLAVS |
| | Probability: 0.997 AA1: | |
| 102379 | 34 AA2: 35 | MKIKLFLLLMIIAISAPMFGLDRVMDFGGLLSSA |
| | Probability: 0.901 AA1: | IYHNSTGTLAISGGTLSASYCAVMNQFIGTVS |
| 102387 | 40 AA2: 41 | ITGGTVSA |
| | Probability: 0.750 AA1: | |
| 102389 | 27 AA2: 28 | LNTILNYIFILGLGPIPGLGVRGAAIA |
| | Probability: 0.996 AA1: | |
| 102399 | 23 AA2: 24 | MNRQHFSFAAALLALVFSLVTCG |
| | Probability: 0.979 AA1: | |
| 102401 | 21 AA2: 22 | MGKRIITGIVLVIFASSVGYG |
| | Probability: 0.986 AA1: | |
| 102418 | 15 AA2: 16 | MRKITLILLFLSVFA |
| | Probability: 0.938 AA1: | |
| 102419 | 26 AA2: 27 | MKKFLPLLVAVLFFSCTSPPPGPRIE |
| | Probability: 0.740 AA1: | |
| 102421 | 17 AA2: 18 | MSSGGAVIISSSGTVNA |
| | Probability: 0.974 AA1: | |
| 102436 | 21 AA2: 22 | MRINFAIIIALIAGLLVLSTG |
| | Probability: 0.874 AA1: | |
| 102438 | 28 AA2: 29 | MAVDVRHPIDRWLLMFALLLAVMGMVWV |
| | Probability: 0.999 AA1: | |
| 102445 | 24 AA2: 25 | GIRKLLHFTVILAVSLFITGLLFA |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 102450 | 21 AA2: 22 | MKAKLFLLFPVMLIISAPLFA |
| | Probability: 0.998 AA1: | |
| 102452 | 25 AA2: 26 | MKTKKLIALLLVFVFSMQHLQFVRA |
| | Probability: 1.000 AA1: | MLNNKKDPVYITKMCMIIVLGALIASLA |
| 102474 | 38 AA2: 39 | TTAFSPRAKA |
| | Probability: 0.633 AA1: | MKAATISKIENYTFNSINTLFLVLLSFITLYPF |
| 102489 | 42 AA2: 43 | WTTIAVSFN |
| | Probability: 0.950 AA1: | |
| 102503 | 25 AA2: 26 | MGTMQTPSNQILISLALAISFFVMA |
| | Probability: 0.617 AA1: | |
| 102504 | 30 AA2: 31 | VPLKLYKGSVSSAGASSPYSLYNASIASFA |
| | Probability: 0.998 AA1: | |
| 102507 | 25 AA2: 26 | MKSGKRDTILTLLALAGFAYFLSWA |
| | Probability: 0.659 AA1: | MNNKVKSIFATFIAHCSPFIPHYYKE |
| 102529 | 58 AA2: 59 | SRRDAVFILVAGLLTLLILAVPTGFKGSAVQA |
| | Probability: 0.991 AA1: | |
| 102560 | 24 AA2: 25 | MSNKRPALGTGLSSLMTQAGHAPA |
| | Probability: 0.520 AA1: | |
| 102605 | 25 AA2: 26 | MTTVQRRRVVITGMGTVNPCGNTVA |
| | Probability: 1.000 AA1: | |
| 102610 | 33 AA2: 34 | MRHKRLGRGPRALFLLTLILFVLLAATAGTGGA |
| | Probability: 0.967 AA1: | |
| 102629 | 23 AA2: 24 | VIIIVLWLAGIAAAMQFAKFSFA |
| | Probability: 0.944 AA1: | |
| 102650 | 25 AA2: 26 | LGTSLKLPLSTTYVTFMVAMGSSLA |
| | Probability: 1.000 AA1: | |
| 102662 | 23 AA2: 24 | MKPRLPALFIFLCFLFVCAGLYA |
| | Probability: 0.994 AA1: | |
| 102672 | 21 AA2: 22 | LKDSRVIIIGGGPAGMAAALA |
| | Probability: 0.873 AA1: | |
| 102680 | 36 AA2: 37 | MERRMLEKIIRYFAQRHMLANFITIAVFLGAVVAWQ |
| | Probability: 0.940 AA1: | |
| 102688 | 19 AA2: 20 | AFGIILAAAAIAFAMTACM |

| | | |
|---|---|---|
| | Probability: 0.999 AA1: | |
| 102716 | 22 AA2: 23 | MKKVTVLAALLSILAASICTYG |
| | Probability: 0.863 AA1: | |
| 102741 | 24 AA2: 25 | MKKKNIYCLLYLFLLVSVFFHLSS |
| | Probability: 0.885 AA1: | |
| 102748 | 20 AA2: 21 | VYKYATGISASLALAQRVAS |
| | Probability: 0.998 AA1: | |
| 102757 | 23 AA2: 24 | MKRTYAIAALAMLALLSACKNST |
| | Probability: 0.993 AA1: | |
| 102765 | 18 AA2: 19 | MRKFLPLLLTALLLIACG |
| | Probability: 0.851 AA1: | |
| 102771 | 30 AA2: 31 | MKIVKKILIILPITIFAIFLILLVTPVLFK |
| | Probability: 1.000 AA1: | |
| 102774 | 23 AA2: 24 | MKKLFIAAAPFFLLILASGNFAG |
| | Probability: 0.885 AA1: | MSACQRKAEPKWGALNLKNIIFAAAAKLLV |
| 102775 | 41 AA2: 42 | ALVLFAGQGMS |
| | Probability: 0.974 AA1: | |
| 102779 | 27 AA2: 28 | LVTFKRLFLVVFAGALMSFNIKTFVHA |
| | Probability: 0.992 AA1: | |
| 102805 | 21 AA2: 22 | MAMGALALAMVGVIALAPVLA |
| | Probability: 0.992 AA1: | |
| 102847 | 23 AA2: 24 | VPFVNDATALIRALAIALPLAFA |
| | Probability: 0.972 AA1: | |
| 102852 | 20 AA2: 21 | VGRLLIAGALVLALNHTVEA |
| | Probability: 0.754 AA1: | |
| 102858 | 32 AA2: 33 | TLLMIMGLAYDSEEGRSVAAALSALLSGEAYA |
| | Probability: 0.932 AA1: | |
| 102859 | 19 AA2: 20 | MVSNSANSGVVTAVFPATA |
| | Probability: 1.000 AA1: | |
| 102862 | 21 AA2: 22 | MKRTVFLALALFALMAVGVYA |
| | Probability: 0.999 AA1: | |
| 102878 | 25 AA2: 26 | NLLRSLLIGTLGSSVLILITVAAMA |
| | Probability: 1.000 AA1: | |
| 102907 | 19 AA2: 20 | MKRAAVVLALLSFTAAAWG |

| | | |
|---|---|---|
| | Probability: 0.718 AA1: | |
| 102914 | 17 AA2: 18 | MATKKFAPMLASNPAFA |
| | Probability: 0.511 AA1: | |
| 102959 | 31 AA2: 32 | VAPIVGIYARADAGSITSTLASASVSAIVCA |
| | Probability: 0.755 AA1: | |
| 102966 | 21 AA2: 22 | MKKPRLLLAGAHSGAGKTSVA |
| | Probability: 0.975 AA1: | |
| 103008 | 18 AA2: 19 | MKKSTMFLLWLGAAISVS |
| | Probability: 0.932 AA1: | |
| 103012 | 21 AA2: 22 | VTFADANSAATTFTMPGNAVA |
| | Probability: 0.552 AA1: | VSIGEVTMLLRGGILQVLTLAAPLLLSALL |
| 103026 | 40 AA2: 41 | VGLIVAILQA |
| | Probability: 0.984 AA1: | |
| 103071 | 18 AA2: 19 | MAVIAGCATLALVGQSVA |
| | Probability: 0.755 AA1: | |
| 103108 | 21 AA2: 22 | MKKPRLLLAGAHSGAGKTSVA |
| | Probability: | FAVLNLFAFLVSWNRKWRVTSFVGLFL |
| | 0.871 AA1: | |
| 103116 | 52 AA2: 53 | NIAGAAYICLFAGPGSFNSPSIAPA |
| | Probability: 0.636 AA1: | |
| 103119 | 20 AA2: 21 | ATGNAASSLAGAAASLGLKS |
| | Probability: 0.716 AA1: | AYYEESENKFHSYKVTRILLGLAFVLCVGFL |
| 103132 | 42 AA2: 43 | GGFFGIGAGWA |
| | Probability: 0.961 AA1: | |
| 103146 | 30 AA2: 31 | MRYAGVMISGALAGLGGLVFIIPTSTQFNA |
| | Probability: 0.975 AA1: | |
| 103184 | 32 AA2: 33 | MINIFINFFKKSIPASFVFSVFSGLIIGTLFG |
| | Probability: 1.000 AA1: | |
| 103235 | 26 AA2: 27 | KMFQTRVIPTLAFCLCAISSIGSAQA |
| | Probability: 0.848 AA1: | MSPLGLLRIAFRALLRNRMRTFLSVLG |
| 103242 | 42 AA2: 43 | IVIGVAAVITMVAIG |
| | Probability: 0.728 AA1: | |
| 103260 | 26 AA2: 27 | MKILVTPASFKQGASAAVMDALKSFA |
| | Probability: 0.974 AA1: | |
| 103264 | 20 AA2: 21 | MNKRVLGADLLLLLTAAVWG |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 103299 | 27 AA2: 28 | MKNKRYRLSQICIAALGAALFACSAGG |
| | Probability: 0.578 AA1: | |
| 103314 | 18 AA2: 19 | ELGRKRIALLGMTFKANN |
| | Probability: 0.857 AA1: | |
| 103354 | 23 AA2: 24 | VDHSLKWRLGLVAAVFMVFTALV |
| | Probability: 0.993 AA1: | |
| 103359 | 23 AA2: 24 | MLTLVVLVLCAVGLSGACSLTEA |
| | Probability: 0.705 AA1: | |
| 103366 | 33 AA2: 34 | GVDPYRVACTGGFVAAGCIGFGAGLTGCTADFA |
| | Probability: 0.864 AA1: | |
| 103381 | 22 AA2: 23 | ECIPARQVTTFAALIGAIPAYA |
| | Probability: 0.692 AA1: | MRTGSAADLMLGLSVFLRRGGYRGGGRRCFALPA |
| 103386 | 47 AA2: 48 | VALLFAAALPAAA |
| | Probability: 0.997 AA1: | |
| 103390 | 28 AA2: 29 | VIKKFVIPVKYQLFLLLLPFLILITAFA |
| | Probability: 0.747 AA1: | MPNHRNLWDASKLKSPFRPCYNLFYSEMTIQSKRAAY |
| 103415 | 56 AA2: 57 | RVLPRFVCLILLAANAVSA |
| | Probability: 0.933 AA1: | |
| 103435 | 32 AA2: 33 | LKIIKGDASFLLLALIILLIVSGIVISIIVFR |
| | Probability: 1.000 AA1: | |
| 103439 | 25 AA2: 26 | MRDAARKYAVFLIALICGLSSSRSA |
| | Probability: 0.584 AA1: | |
| 103440 | 15 AA2: 16 | DRLLFVSLGAVLGGA |
| | Probability: 0.998 AA1: | |
| 103475 | 20 AA2: 21 | MKISFKLMAIMTALGLITIA |
| | Probability: 0.856 AA1: | |
| 103502 | 31 AA2: 32 | KHTSPKHGPNLFFDANRVTALAVVIAIPAAA |
| | Probability: 0.937 AA1: | |
| 103536 | 26 AA2: 27 | MRGFRVRVCLAVLLPVFATITGCAKK |
| | Probability: 1.000 AA1: | |
| 103540 | 23 AA2: 24 | MKNVSKHIGIAVIAAIATAFAAA |
| | Probability: 0.581 AA1: | |
| 103550 | 24 AA2: 25 | MARKSGLSARTLQSIFLDAGVVYA |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 103556 | 21 AA2: 22 | MKRIYFLLVISVISLPPASFA |
| | Probability: 0.930 AA1: | |
| 103576 | 30 AA2: 31 | MKKRIYRDFFVLVLICVICLSVPIIFIFYG |
| | Probability: 0.901 AA1: | |
| 103585 | 19 AA2: 20 | MKLFEIVLIAIGLSMDAFA |
| | Probability: 1.000 AA1: | MAWRNMFRERRRAFIVLLSLFLGITVFTGA |
| 103615 | 40 AA2: 41 | MAIVSSTDIA |
| | Probability: 0.997 AA1: | |
| 103618 | 37 AA2: 38 | MNFKWIWFVAARYIFRKRKKSPAPVLSILGIAVGVLA |
| | Probability: 0.998 AA1: | |
| 103679 | 20 AA2: 21 | MKRKTILFTVVLFIASILPA |
| | Probability: 0.999 AA1: | |
| 103680 | 17 AA2: 18 | LPRLICLILLAAVAVSA |
| | Probability: 0.996 AA1: | |
| 103736 | 15 AA2: 16 | VPFLFLLLAISIAFA |
| | Probability: 0.999 AA1: | |
| 103755 | 24 AA2: 25 | MRGILRFASLSMALWIFLPFALFA |
| | Probability: 1.000 AA1: | |
| 103799 | 37 AA2: 38 | VKFIRFAENCSFNLCLFALLLTGLILAGSAFPPPLHG |
| | Probability: 0.615 AA1: | FVLLAVSRIKRQLRERALLAYFGFGLSALVA |
| 103827 | 44 AA2: 45 | YVFMISLILVVSG |
| | Probability: 0.980 AA1: | |
| 103845 | 28 AA2: 29 | MQTLPQLKKSMSSAVALRSIISSQKSIA |
| | Probability: 0.738 AA1: | MRVTEKLKSYKPKSLDQAIIFLVFVILLVSVGL |
| 103849 | 43 AA2: 44 | FLFFNFTKLS |
| | Probability: 0.530 AA1: | |
| 103857 | 19 AA2: 20 | MKKRIAVLGATGSIGKSSL |
| | Probability: 0.605 AA1: | |
| 103864 | 19 AA2: 20 | MNLATVLTTSRIVIAPVFA |
| | Probability: 0.593 AA1: | |
| 103897 | 32 AA2: 33 | VPKLLFSAFMHLSFLWSICPLGSFDLNTPSWA |
| | Probability: 0.984 AA1: | |
| 103902 | 19 AA2: 20 | MRFLSCVLLVLLLSSCVGR |

| | | |
|---|---|---|
| | Probability: 0.994 AA1: | |
| 103904 | 26 AA2: 27 | MISKLTKFVQMAAIGFALTLTLSCSG |
| | Probability: 1.000 AA1: | |
| 103913 | 23 AA2: 24 | MKKYTILIVCCLISLVAVATVSA |
| | Probability: 0.998 AA1: | |
| 103918 | 21 AA2: 22 | AGKTVTALLAALPALNSGRQA |
| | Probability: 0.937 AA1: | |
| 103927 | 18 AA2: 19 | MILASVAVLSVFTLNRSA |
| | Probability: 0.928 AA1: | |
| 103949 | 21 AA2: 22 | MLNTRLLTVLLLFIITILFFS |
| | Probability: 0.967 AA1: | |
| 103950 | 30 AA2: 31 | MTIVLTTALLSLAIAFILGLALGFFKEFFA |
| | Probability: 0.768 AA1: | |
| 103990 | 20 AA2: 21 | TPYSPILLLIAHCSLLIAQT |
| | Probability: 0.920 AA1: | TYNCRNRFINLTAVGCTAILGFHTFVNIAMT |
| 103998 | 39 AA2: 40 | LGMTPVTG |
| | Probability: 0.751 AA1: | |
| 104005 | 25 AA2: 26 | VITRFTTVTILSKIAAGAGHTLALK |
| | Probability: 0.977 AA1: | |
| 104038 | 22 AA2: 23 | MTIFSKTLLLTALLAFSLTGCK |
| | Probability: 0.984 AA1: | |
| 104052 | 19 AA2: 20 | MKKLIQFIVLCCALSWAIA |
| | Probability: 0.980 AA1: | |
| 104074 | 29 AA2: 30 | VLTLLVLVTGYIGLAMAGGFISQSFSGLA |
| | Probability: 0.994 AA1: | |
| 104076 | 22 AA2: 23 | KKTHFKNATAIILAALFLAISG |
| | Probability: 1.000 AA1: | |
| 104078 | 25 AA2: 26 | MKHLKIIIFLLCVLTLFPIPLKAQS |
| | Probability: 0.927 AA1: | |
| 104080 | 20 AA2: 21 | VNKKYDYLIVGAGLFGAVFA |
| | Probability: 0.532 AA1: | |
| 104094 | 18 AA2: 19 | ITPSRRPSAGASLAGAQG |
| | Probability: 0.990 AA1: | |
| 104104 | 28 AA2: 29 | MNLINTEKITKKLSMALWLAVSFGVVMS |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 104109 | 19 AA2: 20 | MKRLLIILLVLAVAVSAFA |
| | Probability: 0.578 AA1: | |
| 104122 | 20 AA2: 21 | MRETIGLGLIGAGRAGMIHA |
| | Probability: 0.950 AA1: | |
| 104124 | 29 AA2: 30 | MGGIKKEFCPLPSSTTPLTVLGAAARAFA |
| | Probability: 0.963 AA1: | |
| 104143 | 30 AA2: 31 | PGRAVHENWFEYLKAAGIITLFSIANSALA |
| | Probability: 0.999 AA1: | MKKEKHAMKSKITALASRPVKWAVRHKWK |
| 104149 | 44 AA2: 45 | LAVALCALLAVSYFS |
| | Probability: 0.998 AA1: | LNLSLFISRRYLFAKKSHNVINFISLVSAVGVAV |
| 104150 | 47 AA2: 48 | GSLALVVALSVYN |
| | Probability: 0.996 AA1: | |
| 104172 | 26 AA2: 27 | MKQANQDRGTLLLALVAGLSINGTFA |
| | Probability: 1.000 AA1: | |
| 104173 | 20 AA2: 21 | MIKKASLLTACSVTAFSAWA |
| | Probability: 1.000 AA1: | |
| 104174 | 22 AA2: 23 | MKRIAFLILAVFSLLRTNAVFA |
| | Probability: 1.000 AA1: | |
| 104183 | 31 AA2: 32 | MKRFAKYLKTRPIGIVSLVLLLLLYFMMVFA |
| | Probability: 0.919 AA1: | |
| 104185 | 16 AA2: 17 | DQEKLFVLGLQPSAVA |
| | Probability: 0.521 AA1: | |
| 104187 | 17 AA2: 18 | DNYGNLILAGSGTQTVA |
| | Probability: 0.677 AA1: | MSDRVIHTSDTKTLIKNNIRNYSMFIMLALIM |
| 104192 | 41 AA2: 42 | IIFAPLTKG |
| | Probability: 0.779 AA1: | |
| 104204 | 33 AA2: 34 | MKKILLTLIDLIFLTRPVVVVPVWGFCAFGVYA |
| | Probability: 0.999 AA1: | |
| 104211 | 31 AA2: 32 | MSTRIFCRSARRWGGAVWVAVLLCLSFAVGG |
| | Probability: 1.000 AA1: | |
| 104215 | 21 AA2: 22 | MKKVIALSIIMILVCSMAVFA |
| | Probability: 0.998 AA1: | |
| 104220 | 25 AA2: 26 | MKTTRNRLVFAAVFAALIAVSNFFQ |

| | | |
|---|---|---|
| | Probability: 0.989 AA1: | |
| 104241 | 23 AA2: 24 | MKGAIFAVIVLAFAVSCTAPSSN |
| | Probability: 1.000 AA1: | MLKIDAVKLNKKAKRIKLSKKTPLFRFIIRLLLA |
| 104261 | 53 AA2: 54 | ALVAVVLVTLLINAGMILA |
| | Probability: 0.765 AA1: | |
| 104267 | 23 AA2: 24 | MKILKVLPILIIAGQILSCSSNK |
| | Probability: 0.999 AA1: | |
| 104268 | 19 AA2: 20 | MKRKYLTILAALLIPAALA |
| | Probability: 0.908 AA1: | |
| 104269 | 26 AA2: 27 | MGDIKKTNPIMLVLIYLLMFVFAVLA |
| | Probability: 0.999 AA1: | |
| 104288 | 20 AA2: 21 | MKKAILILIVFLAFTTGTSG |
| | Probability: 0.694 AA1: | |
| 104298 | 20 AA2: 21 | ALAILLLLILIAIIIFASRR |
| | Probability: 0.516 AA1: | |
| 104315 | 29 AA2: 30 | ALIISSSEKDTVLYTGFLSAASIHQVASA |
| | Probability: 0.987 AA1: | LNAKDGFLLVFAAWVLASFIGAIPFFLWGINFTD |
| 104331 | 50 AA2: 51 | AFFESACTFATTGATA |
| | Probability: 1.000 AA1: | |
| 104369 | 26 AA2: 27 | MKITTIRTRIAVFIGALALCAAVASA |
| | Probability: 0.952 AA1: | |
| 104411 | 23 AA2: 24 | AEVLVIGGGPAGLCAALAAASLG |
| | Probability: 0.956 AA1: | TTSITVSKATPAALTWPTAAAITYGAALSTSTL |
| 104417 | 45 AA2: 46 | SGGSNTGTFAWT |
| | Probability: 0.984 AA1: | MAKTSQRKTFFRVTLAAFAVSCLLFLTSLDNKF |
| 104425 | 45 AA2: 46 | FDLFLRFLPSLT |
| | Probability: 0.991 AA1: | |
| 104465 | 21 AA2: 22 | MAKKFICVLVVLAVLLAEIVA |
| | Probability: 0.757 AA1: | |
| 104475 | 22 AA2: 23 | LLAIKDCTFGIVGLGLMGGAIA |
| | Probability: 0.527 AA1: | MIVNNGKLSSFITRFSFLNGFKESGKDTVFIAIASFL |
| 104496 | 55 AA2: 56 | TILVMLIPTGFEGSGVYA |
| | Probability: 0.999 AA1: | |
| 104531 | 28 AA2: 29 | MKRKLLIIAAILICGLICATCSILSGLS |

| | | |
|---|---|---|
| | Probability: 0.744 AA1: | |
| 104535 | 33 AA2: 34 | APLLSSEAVPQVAKIALAGFATFAVLPNAGVFA |
| | Probability: 0.989 AA1: | |
| 104565 | 25 AA2: 26 | MKKYFSSLIILLLLVTQITCTGVSG |
| | Probability: 0.992 AA1: | |
| 104577 | 25 AA2: 26 | VNLKASITLFIILLIFSGSPVKIQA |
| | Probability: 0.983 AA1: | |
| 104582 | 24 AA2: 25 | MVFVAKRLVITLLTLFLVSVITFA |
| | Probability: 0.945 AA1: | |
| 104591 | 27 AA2: 28 | MKLLVGFGIMVALILTVSLLSISNLLG |
| | Probability: 0.920 AA1: | |
| 104600 | 22 AA2: 23 | MAKKIIATAIIAIFGIIINFSA |
| | Probability: 0.957 AA1: | MQKVDLKISSMSLANLRAKPVRTVCLFVVVA |
| 104607 | 44 AA2: 45 | ILAFTLFGGSILA |
| | Probability: 0.973 AA1: | |
| 104629 | 31 AA2: 32 | MTEKQKKILLISGGALLVLLITGLILFLALR |
| | Probability: 0.917 AA1: | |
| 104634 | 25 AA2: 26 | PFLYALFGLWLGQAAAAPIISGAAA |
| | Probability: 0.962 AA1: | |
| 104645 | 20 AA2: 21 | MSKFICTSAIDGAVAFVAQA |
| | Probability: 1.000 AA1: | |
| 104665 | 25 AA2: 26 | MKIFGTVKAATVAAAICLAFADVSA |
| | Probability: 0.932 AA1: | |
| 104673 | 35 AA2: 36 | MRPGFARFLAKPNGKFGLLLVGFFFLIAILAPIIA |
| | Probability: 0.757 AA1: | |
| 104687 | 33 AA2: 34 | TSGLYQQHQFGGKTAKPSLAGITLSATGTHTFA |
| | Probability: 1.000 AA1: | |
| 104696 | 21 AA2: 22 | MPTPKLTAALACTLLCTALFA |
| | Probability: 0.806 AA1: | |
| 104743 | 21 AA2: 22 | MAHISNASLVCMGISAVIAIG |
| | Probability: 0.860 AA1: | |
| 104771 | 21 AA2: 22 | MNVARITLVGFGLLAGSIAIA |
| | Probability: 0.999 AA1: | |
| 104791 | 20 AA2: 21 | CSFLAVLGAAGGAFFAGALA |

| | | |
|---|---|---|
| | Probability: 0.997 AA1: | |
| 104802 | 26 AA2: 27 | ILNIFGVSTASLIAILGAAGVTVALA |
| | Probability: 1.000 AA1: | |
| 104841 | 20 AA2: 21 | MRFFVLMAVCAFVLCQRAVA |
| | Probability: 0.991 AA1: | |
| 104858 | 22 AA2: 23 | LPAGIVISATGVISGTPTTAAA |
| | Probability: 0.682 AA1: | |
| 104888 | 19 AA2: 20 | MSLPVVWLCVGIFALVVLA |
| | Probability: 0.644 AA1: | |
| 104903 | 22 AA2: 23 | VKGLAANGSLQITVSATNPATS |
| | Probability: 1.000 AA1: | |
| 104938 | 24 AA2: 25 | VGVSPLYRGLTLLFAVAMAGTVYA |
| | Probability: 0.995 AA1: | MTGGLFELPKKLSFNNYVTVLRGGFYKFFINSILV |
| 104948 | 50 AA2: 51 | LGICLTLLLFFSAFA |
| | Probability: 1.000 AA1: | |
| 104965 | 21 AA2: 22 | MLTLKLTIALACTLLCTTIFA |
| | Probability: 0.988 AA1: | MNKKRLFDWWRRGSAFVWLSGSTLALSLLMVA |
| 104978 | 39 AA2: 40 | GLVLLLA |
| | Probability: 0.995 AA1: | |
| 104983 | 32 AA2: 33 | MKLSVKISLLVGIVLLASTVALLFVVGLKVSG |
| | Probability: 0.689 AA1: | |
| 104985 | 30 AA2: 31 | AAVLALFWGLSAITAWPAGPSGLALGLVLA |
| | Probability: 0.807 AA1: | |
| 105007 | 31 AA2: 32 | MKPVLIATKNPHKAGELAALTAGLAGALALS |
| | Probability: 0.999 AA1: | |
| 105021 | 23 AA2: 24 | MGKFARFGLTLLLTTATIVGAVA |
| | Probability: 0.999 AA1: | |
| 105040 | 25 AA2: 26 | MNKKLFVGITGILIAAALIAGCTLG |
| | Probability: 0.834 AA1: | |
| 105042 | 23 AA2: 24 | MLSNGSSKTALIIGAGPAGLTAA |
| | Probability: 0.949 AA1: | MRTTEKIKLYKPKNLAQSIVLLIFVVLLVSIGLF |
| 105050 | 43 AA2: 44 | FFGFTRLK |
| | Probability: 1.000 AA1: | |
| 105076 | 20 AA2: 21 | MIKKIYYLLILLIISAPALA |

| | | |
|---|---|---|
| | Probability: 0.624 AA1: | |
| 105079 | 31 AA2: 32 | ATVTTTPPNPAGATAAAIAEAKGKAALAAQA |
| | Probability: 0.779 AA1: | |
| 105080 | 36 AA2: 37 | MIRLISIIVVFAIFLCFIMFNLPNKCDISFGFVSFK |
| | Probability: 1.000 AA1: | MTKARVVFHRFLFRRCFRLCCAAAAIVLSAW |
| 105106 | 37 AA2: 38 | MNVAHA |
| | Probability: 0.998 AA1: | |
| 105108 | 31 AA2: 32 | VRQAKIAPAAYRWALALALMLPPLFPSSAPA |
| | Probability: 0.999 AA1: | MATIYSIKRSLVRFNAAQLCVLLLILAIFTIAFIFP |
| 105115 | 46 AA2: 47 | LGMLFSKAFT |
| | Probability: 0.880 AA1: | |
| 105118 | 23 AA2: 24 | MPLVSRQFRHFWLIAVFAVVAQA |
| | Probability: 1.000 AA1: | |
| 105170 | 21 AA2: 22 | MLRSIFLSALLTALSLQNVFA |
| | Probability: 1.000 AA1: | |
| 105212 | 23 AA2: 24 | MKKTIQVLLCAALLLVLGTGCAS |
| | Probability: 0.824 AA1: | |
| 105222 | 15 AA2: 16 | NSGVITAVFSSSATA |
| | Probability: 0.679 AA1: | |
| 105223 | 32 AA2: 33 | ASVSYSEGLTKSYMLDRASVLAALSLTLPASG |
| | Probability: 0.996 AA1: | |
| 105236 | 22 AA2: 23 | MKKTIFYLILISCIAVCGFSQT |
| | Probability: 0.830 AA1: | MAAAFRKNKGRIDVREGFAIVTFSWLATTFIGAIP |
| 105297 | 44 AA2: 45 | FFTFLSAHS |
| | Probability: 0.818 AA1: | |
| 105298 | 27 AA2: 28 | MFISKYRQAIRFKILALTLFTVLTRCA |
| | Probability: 0.885 AA1: | |
| 105313 | 17 AA2: 18 | MKNRKTKIVCSLGPASA |
| | Probability: 0.990 AA1: | |
| 105327 | 21 AA2: 22 | MRLLFWAILTVVAVWAALAWV |
| | Probability: 0.997 AA1: | |
| 105329 | 23 AA2: 24 | MVSRRVLTAMLAMALVLGMTACG |
| | Probability: 1.000 AA1: | |
| 105332 | 24 AA2: 25 | MLRSSFTGIGAAAAGVMRFFSALA |

| | | |
|---|---|---|
| | Probability: 0.908 AA1: | |
| 105341 | 31 AA2: 32 | VRSRNDQVLTDLRLYLRRRLASLALLSARLA |
| | Probability: 1.000 AA1: | |
| 105369 | 36 AA2: 37 | MNNIKMGPKLIASFLFLAALAAFMGIRSISSLKGVA |
| | Probability: 0.947 AA1: | MTPIALMRIAFKALLRNRMRTFLSVLGIVIGVAAV |
| 105373 | 42 AA2: 43 | ITMVAIG |
| | Probability: 1.000 AA1: | |
| 105412 | 22 AA2: 23 | MRSRMFLCMALLSALFASAAVS |
| | Probability: 0.994 AA1: | |
| 105429 | 28 AA2: 29 | MNSLPKRRFLFAVSALFAAVLYILIRYA |
| | Probability: 1.000 AA1: | |
| 105472 | 22 AA2: 23 | MNVSRILATALIVLTAVTTAMA |
| | Probability: 0.989 AA1: | |
| 105486 | 21 AA2: 22 | MTCKRIVPTLLALLLSLYAMV |
| | Probability: 0.996 AA1: | |
| 105521 | 25 AA2: 26 | MINQRKYITNLVLVALFASLISAGA |
| | Probability: 0.970 AA1: | |
| 105557 | 24 AA2: 25 | MSIKGKVAVVTGSATGIGQGVALA |
| | Probability: 0.817 AA1: | |
| 105575 | 22 AA2: 23 | MRKLLFTAFTTIAILGCSNNDM |
| | Probability: 0.670 AA1: | |
| 105596 | 36 AA2: 37 | MDDYTIDKERRMKGNYRFSALIIAALIAIPAGYGQS |
| | Probability: 0.999 AA1: | |
| 105615 | 28 AA2: 29 | MRRGCRQLFILARLTAAAWCVCVFTAAA |
| | Probability: 0.535 AA1: | |
| 105618 | 26 AA2: 27 | MLKQISIKNLALIAGAEISFSNGFWA |
| | Probability: 0.838 AA1: | |
| 105626 | 36 AA2: 37 | MSRFFKKQGFTLIELMVVIVIIGVLASLAIPRFTEA |
| | Probability: 0.987 AA1: | |
| 105635 | 36 AA2: 37 | MRSILKKLLKLITHAQVMSVFLAFAVMVFFSYNFMA |
| | Probability: 0.995 AA1: | |
| 105638 | 25 AA2: 26 | MKKNKIAFFALFIMFLIVSCGKSSV |
| | Probability: 0.993 AA1: | |
| 105656 | 25 AA2: 26 | MKNVFKVLGIIAIVAIIGFSMFACG |

| | | |
|---|---|---|
| | Probability: 0.952 AA1: | |
| 105700 | 21 AA2: 22 | ANSGTGTTGGTSAAGTGATTA |
| | Probability: 0.997 AA1: | MGHCLMRMRRALLTFRPAWARLVAMLA |
| 105701 | 37 AA2: 38 | IASLAFAVLT |
| | Probability: 0.664 AA1: | |
| 105702 | 17 AA2: 18 | ANTSCTVILTAFGFVSG |
| | Probability: 1.000 AA1: | |
| 105707 | 20 AA2: 21 | MKNKIIFFTLLAGAALALYA |
| | Probability: 0.989 AA1: | |
| 105712 | 24 AA2: 25 | MKKGIAFLTLVVFTLFFITNQVFS |
| | Probability: 0.969 AA1: | |
| 105719 | 18 AA2: 19 | MMRAMWSAAAGMTAQAYN |
| | Probability: 1.000 AA1: | |
| 105720 | 34 AA2: 35 | MKKKGFKFVLVFAFLIFSTFWFSACASMGGKSKA |
| | Probability: 0.665 AA1: | |
| 105735 | 17 AA2: 18 | TMPASAVAVTANFEAVA |
| | Probability: 0.733 AA1: | |
| 105749 | 35 AA2: 36 | MREKKISFKIKILIPILAIFIVSVLIISVINYRLL |
| | Probability: 0.990 AA1: | |
| 105764 | 33 AA2: 34 | VENFITRFARYYTPAVVLAAVALAVLPTLFLSG |
| | Probability: 1.000 AA1: | |
| 105772 | 17 AA2: 18 | MTKKNAAFLLALAPALA |
| | Probability: 0.978 AA1: | |
| 105776 | 28 AA2: 29 | MKTDFQFFKLYSLPVVLFLCLYPLKAGA |
| | Probability: 0.980 AA1: | |
| 105806 | 17 AA2: 18 | IGIGLVLALGAAGGVLA |
| | Probability: 1.000 AA1: | |
| 105807 | 24 AA2: 25 | MIRKMNVFVGAALVVAVLCGTLFA |
| | Probability: 0.536 AA1: | |
| 105825 | 25 AA2: 26 | KIKKACILTDVSLDASVSNIETAQA |
| | Probability: 0.979 AA1: | |
| 105843 | 28 AA2: 29 | MREETLAVIAVLALSLGVLLTGMFPSYG |
| | Probability: 0.998 AA1: | |
| 105863 | 18 AA2: 19 | MKVKFAIIGAGPAGLSFA |

| | | |
|---|---|---|
| | Probability: 0.995 AA1: | MIPLYKLEKLPRTQKLRKIAKIFTMAETRLSSLLAGSL |
| 105872 | 50 AA2: 51 | LASGLPASNLQV |
| | Probability: 0.787 AA1: | MEAFSSVLPITAIVLIASVILTPMPSGTILMFLVGAALLII |
| 105878 | 49 AA2: 50 | GMGFFTLG |
| | Probability: 0.992 AA1: | |
| 105881 | 25 AA2: 26 | MKRVLFFLLGIIALGAICVGVAVMR |
| | Probability: 0.999 AA1: | |
| 105885 | 20 AA2: 21 | MKIKFKMSIVVIAIMAVVVA |
| | Probability: 0.985 AA1: | |
| 105886 | 32 AA2: 33 | RDLFFRLNWATLHHFTKFALIASLILAAGAVA |
| | Probability: 1.000 AA1: | |
| 105887 | 23 AA2: 24 | MRRLLVAASLVLFLAGFAWSVLA |
| | Probability: 0.822 AA1: | |
| 105901 | 36 AA2: 37 | MRTNAFTKKDIQKLRDESTFAAAVSGVVNSNANVTA |
| | Probability: 0.970 AA1: | |
| 105903 | 19 AA2: 20 | VASGLILLTLQALAAPASS |
| | Probability: 1.000 AA1: | |
| 105911 | 22 AA2: 23 | VFVKNFRYFLVLFLCFIANAWA |
| | Probability: 1.000 AA1: | |
| 105920 | 23 AA2: 24 | MKKLFSIGLLLGVVALFLFLAHA |
| | Probability: 1.000 AA1: | |
| 105946 | 22 AA2: 23 | MRLTNIPVALAALLAVAAPAWA |
| | Probability: 1.000 AA1: | |
| 105947 | 26 AA2: 27 | MVRHRVALSIASMMLLAICLGVFVNA |
| | Probability: 1.000 AA1: | |
| 105953 | 23 AA2: 24 | MRSRMFLWIALLSAVLVSASASA |
| | Probability: 0.810 AA1: | |
| 105954 | 26 AA2: 27 | MKRRRGFTLIEILIALTILSILATLG |
| | Probability: 1.000 AA1: | |
| 105978 | 21 AA2: 22 | MGRLKLLALAGLMAAAFSVSA |
| | Probability: 0.991 AA1: | |
| 105981 | 23 AA2: 24 | MVSPLAACGMGFTFLCQSQLAVA |
| | Probability: 0.748 AA1: | |
| 106000 | 21 AA2: 22 | MTASLVTQVAIGGLLMGLVYA |

| | | |
|---|---|---|
| | Probability: | |
| | 0.998 AA1: | MRHTRGERFWRQLKRDKLTLGACAFLVLVI |
| 106016 | 38 AA2: 39 | AASMLAFL |
| | Probability: | |
| | 0.844 AA1: | |
| 106024 | 31 AA2: 32 | MTHQITRRDFICKTTVGTAALALTPVSGILA |
| | Probability: | |
| | 0.962 AA1: | MMETMNTIAGFMKMLLIAIASVSLLVGSIGIMN |
| 106034 | 40 AA2: 41 | MMLVSVA |
| | Probability: | |
| | 0.722 AA1: | |
| 106039 | 21 AA2: 22 | MRNNLFKLIFLAGVYASLLFS |
| | Probability: | |
| | 1.000 AA1: | |
| 106043 | 23 AA2: 24 | MKKTLIAAAPFFLLIVALGGVVG |
| | Probability: | |
| | 0.722 AA1: | MSEEKTTNTKADKISRPIGAKLIIIITILLLFSLGAITV |
| 106046 | 46 AA2: 47 | LVSVLVS |
| | Probability: | |
| | 0.504 AA1: | MADTYFVSSLGTSQVAAVGIAFPLMAIIQAMGFFFG |
| 106049 | 44 AA2: 45 | QGSGNYMA |
| | Probability: | |
| | 0.997 AA1: | |
| 106051 | 18 AA2: 19 | MFKHFLTIVAFAGVFVFA |
| | Probability: | |
| | 0.687 AA1: | |
| 106081 | 38 AA2: 39 | MLILMAQTPLITLLLGFVYTLSISLFPMSFYFCLAVTG |
| | Probability: | |
| | 1.000 AA1: | |
| 106087 | 39 AA2: 40 | MKNVNNKCAKILRILATQIRRICAIALVAVIALSFALAA |
| | Probability: | |
| | 1.000 AA1: | |
| 106089 | 20 AA2: 21 | MRMILITAVFAAVSVGSAAA |
| | Probability: | |
| | 0.997 AA1: | |
| 106114 | 19 AA2: 20 | MLMGAFVFSFVSTAFNARA |
| | Probability: | |
| | 1.000 AA1: | |
| 106127 | 23 AA2: 24 | MRKRNLFFCFILVLAALPVSVFG |
| | Probability: | |
| | 1.000 AA1: | |
| 106141 | 21 AA2: 22 | MKKIIWVIFAVFAALSLSAAA |
| | Probability: | |
| | 0.520 AA1: | |
| 106166 | 17 AA2: 18 | MNGLLLLLISGITLLLA |
| | Probability: | |
| | 0.967 AA1: | |
| 106169 | 20 AA2: 21 | NRIEIFFMTGTAPDPAAANA |
| | Probability: | |
| | 0.955 AA1: | |
| 106183 | 24 AA2: 25 | DNVTVTATAASCYTFIGWSGAATG |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 106187 | 25 AA2: 26 | MITKRFLLLIGGVLFAMAFIFSCSS |
| | Probability: 0.992 AA1: | |
| 106210 | 20 AA2: 21 | VHKKLGNVSGALAVAFLAGS |
| | Probability: 0.766 AA1: | |
| 106225 | 35 AA2: 36 | TELMATYMYKMSFIRFNMGYGAAVAAGMFILITLA |
| | Probability: 0.578 AA1: | |
| 106246 | 20 AA2: 21 | AGNIYAALRLPLTPPFSTEA |
| | Probability: 0.732 AA1: | |
| 106257 | 19 AA2: 20 | SLSKLLDIGCAYGAFLAAA |
| | Probability: 0.976 AA1: | MLFWKANWIYGVFALASLPFLAVFFIIYFK |
| 106277 | 42 AA2: 43 | KFSKTIPGSVNA |
| | Probability: 0.707 AA1: | |
| 106314 | 23 AA2: 24 | AVCRKTVTVSTAICVLIVAGFWA |
| | Probability: 1.000 AA1: | |
| 106332 | 28 AA2: 29 | MNRRKMVCLQALLLLIFAAWLTPGAAWA |
| | Probability: 0.629 AA1: | |
| 106363 | 18 AA2: 19 | MTDNNIGPVLLGAVIANA |
| | Probability: 0.997 AA1: | |
| 106370 | 23 AA2: 24 | MNMKRNIVFPVVLAALATAMVLS |
| | Probability: 0.997 AA1: | |
| 106373 | 36 AA2: 37 | MNISEPFIRRPIATALLMAAIGFVGLVAIPFLPVAA |
| | Probability: 1.000 AA1: | |
| 106381 | 23 AA2: 24 | MKRTFFAAAAFFAVLAAASGAYA |
| | Probability: 0.912 AA1: | |
| 106385 | 26 AA2: 27 | VGPARQVREAALSGAVALLAAGRVLA |
| | Probability: 0.991 AA1: | |
| 106397 | 29 AA2: 30 | MKQTSLRTFGPVAWAALVLLGFSSGLPLA |
| | Probability: 0.996 AA1: | |
| 106408 | 21 AA2: 22 | MNKFVRGGLTLLLAAGLAVVG |
| | Probability: 0.996 AA1: | |
| 106433 | 22 AA2: 23 | GRFAATTFFIIFALLIGLGVFA |
| | Probability: 1.000 AA1: | |
| 106442 | 22 AA2: 23 | MKRKFFHIIPVLFVFVSVSAFA |

| | | |
|---|---|---|
| | Probability: 0.982 AA1: | |
| 106447 | 22 AA2: 23 | MKISLWKAMCTGLLCAAVLLTA |
| | Probability: 0.986 AA1: | |
| 106458 | 21 AA2: 22 | WVWSFFGCGSLASVTIGNSVA |
| | Probability: 0.932 AA1: | |
| 106483 | 16 AA2: 17 | MKKHLFFLSILAACYG |
| | Probability: 0.957 AA1: | |
| 106503 | 18 AA2: 19 | IFPGWAVSASASAFILVA |
| | Probability: 0.990 AA1: | |
| 106508 | 26 AA2: 27 | MKRLFLFPAWIFILSLTIFSGGCRQA |
| | Probability: 0.968 AA1: | |
| 106540 | 21 AA2: 22 | MLLGVFSVLLSATLCLGASLA |
| | Probability: 0.981 AA1: | |
| 106543 | 19 AA2: 20 | MKKYSIIVITLVMLALLGA |
| | Probability: 0.748 AA1: | |
| 106551 | 26 AA2: 27 | MDSRAWLLSHRALLAMLIFLSGQVLS |
| | Probability: 0.996 AA1: | |
| 106559 | 33 AA2: 34 | VKQNKLFIILAAAAIILGILNSSLIMKPPVYNA |
| | Probability: 0.994 AA1: | |
| 106563 | 19 AA2: 20 | MPWRICVALAVFVLSLVYA |
| | Probability: 0.870 AA1: | |
| 106627 | 18 AA2: 19 | TTGKITVSGTAKITSANA |
| | Probability: 0.686 AA1: | |
| 106628 | 35 AA2: 36 | YTYGSTVFYHNSATGTLTISGGLISISTGNSSSYA |
| | Probability: 1.000 AA1: | |
| 106631 | 25 AA2: 26 | MKTSRFLSLAAAFSLFSLFSCTSDV |
| | Probability: 0.985 AA1: | MTRNFTISLKNRLSLTYVLFICLVMLVLCIVINLFT |
| 106674 | 44 AA2: 45 | NIFFSALV |
| | Probability: 0.998 AA1: | |
| 106680 | 29 AA2: 30 | CLDEAAACVLIVIAAFCVLLLPARTSAAA |
| | Probability: 1.000 AA1: | |
| 106697 | 22 AA2: 23 | VTRIAINTAILSILFTALAATA |
| | Probability: 0.980 AA1: | |
| 106699 | 26 AA2: 27 | MSTANNKYDIIVVGSGPAGASAATSA |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 106702 | 19 AA2: 20 | MKKLIVLTMLLAGAGLLNA |
| | Probability: 0.975 AA1: | |
| 106742 | 24 AA2: 25 | MKNTRNAFLNLMVLAGLCYFLGWA |
| | Probability: 0.957 AA1: | |
| 106772 | 20 AA2: 21 | MDATTAAVLAIGLLVFVAHG |
| | Probability: 0.992 AA1: | |
| 106783 | 27 AA2: 28 | MRTFCVFVFALAIMFSMDIFMSLVSLV |
| | Probability: 0.998 AA1: | |
| 106797 | 18 AA2: 19 | MKRGVILLCLVLFVTDFA |
| | Probability: 0.987 AA1: | MSKLNSLMLYLQKQQKGQQFSLIVTLVLVYAFMAFMN |
| 106801 | 51 AA2: 52 | GSRFLSLYNLSSMA |
| | Probability: 0.694 AA1: | |
| 106811 | 27 AA2: 28 | MHNAKFIVPNAFTSLNFLLGVASICWS |
| | Probability: 0.796 AA1: | |
| 106819 | 32 AA2: 33 | MKYLKVRKNRVLISVIITLIVYGILWIVTPIA |
| | Probability: 0.964 AA1: | |
| 106826 | 24 AA2: 25 | MGILVSAIVGIVLTLPISSAALCA |
| | Probability: 1.000 AA1: | |
| 106833 | 25 AA2: 26 | VKKHSIFRPLAAALLLLPSLLAAQS |
| | Probability: 0.814 AA1: | |
| 106866 | 22 AA2: 23 | MPFENSAFPVLTVLTFWPTAAA |
| | Probability: 0.665 AA1: | |
| 106878 | 38 AA2: 39 | NAWGVFGAAFGPAILLSLFWKRFTFAGAVAGIIAGAVA |
| | Probability: 0.612 AA1: | |
| 106897 | 18 AA2: 19 | VFALSRAALVVLGGADSA |
| | Probability: 0.997 AA1: | |
| 106899 | 28 AA2: 29 | MFLLGKIFTWLILSPGLFLVLLFLVLHA |
| | Probability: 0.571 AA1: | MDLTVLIAIGAGIAVVTGLGAGIGIGIATSGFLQAISRQPE |
| 106915 | 57 AA2: 58 | ATGKMTPMFFVGVALA |
| | Probability: 0.991 AA1: | |
| 106938 | 22 AA2: 23 | MSADIFSGILLSLAPGALLANA |
| | Probability: 0.654 AA1: | |
| 106970 | 22 AA2: 23 | MIQQIKKSFVIGCIAMIQPLLM |

| | | |
|---|---|---|
| | Probability: 0.780 AA1: | |
| 106973 | 25 AA2: 26 | ARFLKLFKHGFDFSILALGIALILA |
| | Probability: 0.971 AA1: | |
| 106979 | 23 AA2: 24 | MYKYILALLCLSFMMACKNNNVA |
| | Probability: 1.000 AA1: | |
| 106987 | 20 AA2: 21 | MKLINILAFLLIFFSQSVTA |
| | Probability: 1.000 AA1: | |
| 106996 | 20 AA2: 21 | MKKFVCVFILLFFVSVLAFS |
| | Probability: 0.996 AA1: | |
| 107022 | 24 AA2: 25 | MKKLIAVIMCFCLIMLISCGGGKA |
| | Probability: 0.751 AA1: | |
| 107036 | 27 AA2: 28 | MDEDNSKIRILIIAAVIIAIVALLIFA |
| | Probability: 0.976 AA1: | |
| 107082 | 28 AA2: 29 | MKREKSNIKGSQLLAFFPFLSSFFSCIA |
| | Probability: 0.994 AA1: | |
| 107091 | 19 AA2: 20 | MKKITSIVFVLGLALLASG |
| | Probability: 0.998 AA1: | |
| 107097 | 36 AA2: 37 | MKFNDILKIDYLLLLPALVLLTMGILFIYSSGVTSA |
| | Probability: 0.916 AA1: | |
| 107114 | 19 AA2: 20 | MFPLSLPVLATIGLFLSFA |
| | Probability: 0.999 AA1: | |
| 107131 | 31 AA2: 32 | MKQRNQKTITILTAAFLLAAAVIVTFSCVWE |
| | Probability: 0.623 AA1: | |
| 107147 | 34 AA2: 35 | MASVQKRTAFDFVILCFMLLVIVITLYPFLNVTA |
| | Probability: 0.910 AA1: | TLKGLPILAGICLIFALILIKGTNRAKPAGPILVLA |
| 107149 | 43 AA2: 44 | LFFTALA |
| | Probability: 0.525 AA1: | |
| 107165 | 35 AA2: 36 | MVGMKVIAVFYITAMKGILLAGGAGSRLYPLSAVA |
| | Probability: 0.961 AA1: | |
| 107184 | 19 AA2: 20 | VRGSLVWAFALVMVQNVLS |
| | Probability: 0.997 AA1: | |
| 107190 | 34 AA2: 35 | MRNIKIGPKLIGSSLITAALIAFMGIYSMKSVTA |
| | Probability: 1.000 AA1: | |
| 107203 | 19 AA2: 20 | VRYYSLFLAIFLAVCVAYA |

| | | |
|---|---|---|
| | Probability: 0.729 AA1: | |
| 107206 | 27 AA2: 28 | MAQVYGWMTVGLLLTAFVAWYAANSAA |
| | Probability: 0.964 AA1: | |
| 107217 | 27 AA2: 28 | MKKRTGIFNMAATIVVAMVMFTVDCKA |
| | Probability: 0.598 AA1: | |
| 107240 | 19 AA2: 20 | MDIAKNAMILAAVAALDLA |
| | Probability: 0.966 AA1: | NASNQNVTWTNSNTAAATISATSGGTITVTPVA |
| 107241 | 44 AA2: 45 | VGTTNITAASA |
| | Probability: 0.994 AA1: | |
| 107257 | 34 AA2: 35 | LLLIIGIIAAVLLLILLIFFISRRLGSSPNRVMA |
| | Probability: 0.994 AA1: | |
| 107264 | 21 AA2: 22 | MKKVLTIIVVIFVVFIGIVVA |
| | Probability: 0.839 AA1: | |
| 107268 | 28 AA2: 29 | LNKREFILLLAGVSAVILIIVLCGIIFT |
| | Probability: 0.976 AA1: | |
| 107337 | 16 AA2: 17 | ILTLIAGILLSALTTA |
| | Probability: 0.997 AA1: | |
| 107359 | 23 AA2: 24 | MKNKFNVILAAVFAIILPLSFLS |
| | Probability: 0.999 AA1: | |
| 107394 | 23 AA2: 24 | MKKTVICFSVLYSVLIVIPLAHA |
| | Probability: 0.933 AA1: | |
| 107395 | 28 AA2: 29 | TSHLLAKTILTSVILLALFFALDTVLFG |
| | Probability: 0.999 AA1: | |
| 107397 | 39 AA2: 40 | MRNTIKFRFAGIGIMLAAIAVFSVVVMLLWNALMPQIFA |
| | Probability: 0.928 AA1: | |
| 107432 | 34 AA2: 35 | MPGRRKILLGLYILEKGLEVLCCLLFALPVAVMA |
| | Probability: 0.592 AA1: | |
| 107436 | 35 AA2: 36 | MTMPWRPGRFDWKREIKASGLAALWLTLLTGPLMV |
| | Probability: 0.937 AA1: | |
| 107445 | 28 AA2: 29 | MSNRTLVGRSPSIAPCSPWLIFSNNAFA |
| | Probability: 0.997 AA1: | MRLGLRRRRRRRGRLRRRHGRRRRFRRGRRGRR |
| 107448 | 73 AA2: 74 | RRRRLYRRRLYRHLFLRLRPRRRHLHHLKSLGTLDFLGLN |
| | Probability: 1.000 AA1: | |
| 107451 | 27 AA2: 28 | MVAIPKRMSILALAALLALLAQPSAVA |

| | | |
|---|---|---|
| | Probability: 0.986 AA1: | |
| 107459 | 35 AA2: 36 | MNLNKLRLKWKIFAFLLGFCALLLIVLWLFQIVFL |
| | Probability: 1.000 AA1: | |
| 107492 | 32 AA2: 33 | MTTESKRKIHSWWRPLLLSLALWLGAAFFVGA |
| | Probability: 0.926 AA1: | |
| 107527 | 21 AA2: 22 | MNMVKTAAVAAARFVVTKILA |
| | Probability: 1.000 AA1: | |
| 107531 | 23 AA2: 24 | MKKTALLLTIALAFQVASLYAQA |
| | Probability: 0.882 AA1: | |
| 107544 | 30 AA2: 31 | MHSSMALASSLSPLTCLCRRSNFLPRLSMS |
| | Probability: 0.996 AA1: | |
| 107567 | 40 AA2: 41 | MLLRNSLERFTTVFPFQLITRRFSFVTSAIITASRFSASA |
| | Probability: 0.736 AA1: | MWFXFCVLGFGGFLLGFXXVGVFFLVGVFGVLG |
| 107581 | 67 AA2: 68 | LGGLFVFGVVFLFXXFXGFGFWGWXGXWGVLVLX |
| | Probability: 0.989 AA1: | MVSDTNNAFAKRGAPLVSRKSLDISTMVLPRQFMRRRL |
| 107609 | 65 AA2: 66 | SSLTLAMTTASRFSSAAAFINASASQA |
| | Probability: 0.605 AA1: | |
| 107621 | 23 AA2: 24 | MSLKPSITPIFIFTVILSFPAKM |
| | Probability: 0.844 AA1: | |
| 107628 | 30 AA2: 31 | MKFIHLSVGYALLGFACGTFGLLFNTIIGA |
| | Probability: 0.935 AA1: | |
| 107643 | 20 AA2: 21 | MTALSGISALTAFFFARALA |
| | Probability: 0.602 AA1: | |
| 107645 | 30 AA2: 31 | MQWFSNANGTTAATTPTGITASVSNVSNNA |
| | Probability: 0.977 AA1: | |
| 107663 | 39 AA2: 40 | MIITLLLGEYTMRRKFTKLAQMATLGLALVFTFSCTSGG |
| | Probability: 0.509 AA1: | |
| 107672 | 34 AA2: 35 | MEYINWSEKKIPVTGSYDLFIAGGGCAGAGAAIA |
| | Probability: 0.998 AA1: | MIPLYKLEKLPRTQKLRKIAKIFTMAETRLS |
| 107693 | 45 AA2: 46 | SLLASGLSASSLTA |
| | Probability: 0.993 AA1: | |
| 107698 | 33 AA2: 34 | MIFFFFFKKKKKKTRGGGGVFFLTGGCFFFIFG |
| | Probability: 0.756 AA1: | |
| 107714 | 21 AA2: 22 | MSMPTAATGICLGICTIESSA |

| | | |
|---|---|---|
| | Probability: 1.000 AA1: | |
| 107725 | 22 AA2: 23 | MNRGFFDMAFLVLLLLFGYSLA |
| | Probability: 0.944 AA1: | MIEKTDKKLSLPSRVHPAVIAVWAALVSASHI |
| 107734 | 48 AA2: 49 | PSVPMLGTGSNFSFA |
| | Probability: 0.976 AA1: | |
| 107751 | 25 AA2: 26 | MRNRSRKLYMAHAFAAYLKMGNLNA |
| | Probability: 0.845 AA1: | MSVTGATCGTVTGVSGSNIACGNITIQAASSSSV |
| 107765 | 40 AA2: 41 | APSSSS |
| | Probability: 0.514 AA1: | |
| 107766 | 20 AA2: 21 | MNKLIVIIGFLSVLFLSCPK |
| | Probability: 0.918 AA1: | MQTLPLMATIPRPILFLIIRSNLNTTSESRSSVSF |
| 107775 | 72 AA2: 73 | FSFFNRVSTCFFFSVTSSSMPAFFSASSFSCCSMRAA |
| | Probability: 0.986 AA1: | MSKKALVISTIFCPRQFITRRLSPVMRATTDAS |
| 107784 | 50 AA2: 51 | IFSAAACCSRAFRSVSA |
| | Probability: 0.950 AA1: | MNNVFRALKLDFCITKTSYSKLIMIYSISILLGLS |
| 107785 | 58 AA2: 59 | TQPIMPVIMIMFFCVSFSGLAFS |
| | Probability: 0.994 AA1: | MALTYKKKVLLLSSLVAVLAVIYILTFVL |
| 107866 | 40 AA2: 41 | DPESRSSAAFA |
| | Probability: 0.831 AA1: | |
| 107878 | 19 AA2: 20 | MKSFFIFIHLLFLPIFCFA |
| | Probability: 0.970 AA1: | |
| 107882 | 22 AA2: 23 | MTVASRFSAAASFIKAGASSLA |
| | Probability: 0.997 AA1: | |
| 107887 | 21 AA2: 22 | MFLRIFFSIFLLFSKSTVSWA |
| | Probability: 0.937 AA1: | |
| 107888 | 25 AA2: 26 | MIRLARGPXAGALCSAAGGFVGSLV |
| | Probability: 0.659 AA1: | MIVRIVRLLKLTAVGTSIKSERISTMSADSIATS |
| 107898 | 49 AA2: 50 | VPLPMAIPMSACASA |
| | Probability: 0.831 AA1: | |
| 107900 | 21 AA2: 22 | MPVLCSTVSTSFSLTTPAILA |
| | Probability: 0.858 AA1: | |
| 107908 | 25 AA2: 26 | MRAARRYRYRSRLVSAISVTLRAAA |
| | Probability: 0.512 AA1: | MPVCFLHFPIKFFFLLLHFLFKSFVVFLFFPLI |
| 107914 | 72 AA2: 73 | SFVVFLFFLFKSFVAFLFFPLKPFVVFLFFTLKPFVAFL |

| | | |
|---|---|---|
| | Probability: 0.581 AA1: | |
| 107932 | 22 AA2: 23 | MNMVNTAAVAAARFVVTKILLA |
| | Probability: 0.984 AA1: | |
| 107937 | 24 AA2: 25 | MPKNKFTLIIPALVFFLCVTSCTK |
| | Probability: 0.650 AA1: | |
| 107955 | 27 AA2: 28 | MVPSFEVTVMVALPGLIALISPSSPMV |
| | Probability: 0.998 AA1: | |
| 107956 | 23 AA2: 24 | MKLKPSITAIFIFFFIALTAVYS |
| | Probability: 0.993 AA1: | |
| 107967 | 37 AA2: 38 | MKLFHSLRFRFAFIFSVFIIILSVSLSLLGIRQTTKA |
| | Probability: 1.000 AA1: | |
| 107971 | 20 AA2: 21 | MKKIALVLVVFLMAATLVFA |
| | Probability: 1.000 AA1: | MIAIALTSRRFRWTTYSFILPAIALTGARLSVIFTHVTIS |
| 107975 | 59 AA2: 60 | AVAFSLSAVTAGFFLTALA |
| | Probability: 0.728 AA1: | MTTNSFAWRSPIATAVVWSSTRSSMYMFPSTFPKT |
| 107980 | 44 AA2: 45 | NAFCALISA |
| | Probability: 0.614 AA1: | |
| 107986 | 22 AA2: 23 | MRAYISFFLLPSFLFFLLLRLS |
| | Probability: 0.589 AA1: | MPNCLFVSFISRSNSSFCFFISCSNRSSCSFFSRSYRS |
| 107987 | 71 AA2: 72 | SCSFFSCSNRSSRSFFSRSSRSSCSFFSRSSRS |
| | Probability: 0.994 AA1: | MLLRNSLERFTIVFPFHFITRRFSRVTSAITTASRFSASA |
| 107995 | 53 AA2: 54 | SSINFFASFSAVT |
| | Probability: 0.996 AA1: | |
| 108008 | 27 AA2: 28 | MHINIRRLTMKRLLFAISIISLCLLFS |
| | Probability: 0.527 AA1: | |
| 108015 | 23 AA2: 24 | MAXRILSYGFSALXVYAPESAYA |
| | Probability: 0.796 AA1: | |
| 108024 | 24 AA2: 25 | MTDSRVFRRYAGGLFFLAAAVFCA |
| | Probability: 0.605 AA1: | |
| 108058 | 25 AA2: 26 | MPFSISAFRSLLVAATRRTSVSTVS |
| | Probability: 0.999 AA1: | |
| 108059 | 20 AA2: 21 | MRRVILAILGLLFTTGFCAA |
| | Probability: 1.000 AA1: | |
| 108066 | 28 AA2: 29 | MLIVTRSIFWPLCITLASTVCLMRSAMA |

| | | |
|---|---|---|
| | Probability: 0.993 AA1: | |
| 108076 | 22 AA2: 23 | MSASSAIFLSGSTIVAAPNLNA |
| | Probability: 0.990 AA1: | |
| 108080 | 22 AA2: 23 | MKLKPSITAFLVFIAVQTAVYS |
| | Probability: | MLIVSITSIQIMPIPLSILIIPLVLIMSIPLSILI |
| | 0.575 AA1: | MPLVLIMSIPMLISMMSLVLM |
| 108083 | 56 AA2: 57 | |
| | Probability: 0.548 AA1: | MVYFVPFFKAAQNRNCIFHGRFRRIHLLKSPFKRSVL |
| 108108 | 52 AA2: 53 | LDMFSVLVQSSRAYT |
| | Probability: 0.990 AA1: | MMSGTNKTFAVRGAPLVSRKFRDM |
| 108120 | 54 AA2: 55 | STMVLPRQLMRRRLSLFTLAMITASRFSVA |
| | Probability: 0.792 AA1: | |
| 108127 | 20 AA2: 21 | MRLNPGYYTPALITHLMAQA |
| | Probability: 0.986 AA1: | |
| 108134 | 25 AA2: 26 | MNLKPSITLTLILALIISFHAEIHA |
| | Probability: 1.000 AA1: | |
| 108144 | 18 AA2: 19 | MLPPALALLLAASPAAHA |
| | Probability: 0.922 AA1: | |
| 108145 | 26 AA2: 27 | MINPAMRIIVVLIMIMLAAGCAGRRA |
| | Probability: 0.994 AA1: | |
| 108154 | 19 AA2: 20 | MQIAFLIFLSFCSCLYAQQ |
| | Probability: 0.990 AA1: | |
| 108173 | 25 AA2: 26 | MFSRLGPAGILMAAALGFTAGLVFN |
| | Probability: 0.987 AA1: | |
| 108185 | 23 AA2: 24 | MKNICLLLLIPLFAACLTNISEQ |
| | Probability: 1.000 AA1: | |
| 108190 | 22 AA2: 23 | MGKKPIILSLFLFFMHSIYAYA |
| | Probability: 0.528 AA1: | MLTTLFLCYLLFVTCYLLSVICYLLFSAFLLKMR |
| 108206 | 47 AA2: 48 | EFHTETCPSLAYA |
| | Probability: 0.924 AA1: | |
| 108215 | 24 AA2: 25 | MSISRNMAFSVWQCVLAIFSCAAA |
| | Probability: 0.892 AA1: | |
| 108227 | 22 AA2: 23 | MNMVKTAAVAAARLVVTKILLA |
| | Probability: 0.917 AA1: | MMSMPSASSPTATETPPAPKSLHRFIRRLT |
| 108234 | 56 AA2: 57 | SGLRKRRWSFRSSTAFPFCTSAPALS |

| | | |
|---|---|---|
| | Probability: 0.998 AA1: | |
| 108240 | 20 AA2: 21 | MTEWPSSAAAAFMKASASQA |
| | Probability: 0.915 AA1: | MSSSSSKASTNFIILTAAPASKLTGVSTTFVISA |
| 108255 | 74 AA2: 75 | EVGSILAFTIASRTASKASTDAVISYSSPSADMSSAPASS |
| | Probability: 0.998 AA1: | |
| 108265 | 31 AA2: 32 | MKSVCKKIAIFLMLVFLACTPGFFLFGQYNA |
| | Probability: 0.942 AA1: | |
| 108276 | 21 AA2: 22 | MKNLVLLSFILNLYNVFSLSA |
| | Probability: 0.879 AA1: | |
| 108277 | 30 AA2: 31 | MLPVCLEAASSSCSLCCSSIALRMFSSCSA |
| | Probability: 0.999 AA1: | |
| 108293 | 28 AA2: 29 | MAYKHFFNQTFTFLALVLIITLSLSACS |
| | Probability: 0.988 AA1: | |
| 108300 | 20 AA2: 21 | MRSLFLIILFAFFAPIKSYA |
| | Probability: 0.972 AA1: | |
| 108302 | 29 AA2: 30 | MVITVRRIAWVLTILFLLFAFASCGGRGA |
| | Probability: 0.722 AA1: | |
| 108304 | 22 AA2: 23 | MNMVKTAAVAAARFVVTKMLLA |
| | Probability: 0.994 AA1: | |
| 108355 | 30 AA2: 31 | MFFVMKHSQKLIILELFPLLFLFSCSSSSG |
| | Probability: 0.924 AA1: | |
| 108364 | 21 AA2: 22 | MIRIIVFILVFVFFLAFIVLN |
| | Probability: 0.708 AA1: | |
| 108432 | 26 AA2: 27 | MARVSRSARASSILMVSISDSGSILS |
| | Probability: 0.507 AA1: | |
| 108437 | 26 AA2: 27 | MKKGFFFFFKFPGGVFFFFPPPFFWG |
| | Probability: 0.842 AA1: | MSSAYAQSLSGDSKYLIWSSASLRRSWMSPNT |
| 108453 | 74 AA2: 75 | LWRSSSSSRLTVIVSFLRCSCWMSCLASSSSIVFDCSSASLA |
| | Probability: 0.980 AA1: | MVMSGAVIMMIVVMAVMIVISIMMIIIVMMIAAVMTV |
| 108461 | 70 AA2: 71 | MVMIVIMEMIIAVMMFVNMVMIIAFDMPINTNN |
| | Probability: 0.997 AA1: | |
| 108470 | 21 AA2: 22 | MRFFTTSLIIAVALFSVSCQS |
| | Probability: 0.990 AA1: | |
| 108484 | 23 AA2: 24 | MIRVSTSTTAPASASAWQTGATA |

| | | |
|---|---|---|
| | Probability: 0.962 AA1: | |
| 108495 | 20 AA2: 21 | MLITSPGAVSAFLTGSELFA |
| | Probability: 0.984 AA1: | |
| 108499 | 34 AA2: 35 | MKITKKLKFLKKINPFRANIILIILGISISLYFT |
| | Probability: 0.984 AA1: | MKKSNFKQNKMKIFKMLGLYALICILEFALLSFFIHIRLF |
| 108501 | 61 AA2: 62 | SGVLYFFRLFIYAILSGSLLL |
| | Probability: 0.912 AA1: | |
| 108518 | 24 AA2: 25 | MPPTLVAFAAGTAPVASIRSGALS |
| | Probability: 1.000 AA1: | |
| 108530 | 39 AA2: 40 | MPRQFMRRRLSFSSLTLAMITASRFSAAAAFIKASASQA |
| | Probability: 0.809 AA1: | |
| 108535 | 37 AA2: 38 | MKSLPFAVQLSLIVFLILIIPFIVLMSYIGSSTLRYS |
| | Probability: 0.996 AA1: | |
| 108546 | 19 AA2: 20 | MKKLIILTLIFFSVMGVFG |
| | Probability: 0.995 AA1: | |
| 108571 | 21 AA2: 22 | MNNIKIGPKLIVSFLIVVALA |
| | Probability: 0.748 AA1: | |
| 108588 | 28 AA2: 29 | MSGSSSTAMMRTLLQFTSGASSPCFSFA |
| | Probability: 1.000 AA1: | |
| 108598 | 22 AA2: 23 | MSFSITLFSMSAFIAASFSARA |
| | Probability: 0.771 AA1: | |
| 108600 | 25 AA2: 26 | MVDNNSGSIAAVVGMVNFAALNVNA |
| | Probability: 0.984 AA1: | |
| 108620 | 25 AA2: 26 | MLIEEKSLMPAAFSFSVFNAPVALA |
| | Probability: 0.503 AA1: | |
| 108634 | 34 AA2: 35 | MKSLIERCSRIFSFVKALPLSSKTAMPASTSCSA |
| | Probability: 0.540 AA1: | |
| 108637 | 28 AA2: 29 | MTGIGRAWRSASGGAFSMAISTDGSLWA |
| | Probability: 0.980 AA1: | |
| 108643 | 22 AA2: 23 | MQTRPPCFCLICGFCSSARDFA |
| | Probability: 0.989 AA1: | MQVIRGVIIALVLLPMRKVFTEERYGFLMLGLLIL |
| 108644 | 44 AA2: 45 | GLSAASTFA |
| | Probability: 0.980 AA1: | |
| 108651 | 34 AA2: 35 | MKQWSNPRKYSRLFIVFLNNRGIVLLLGLLGRFA |

| | | |
|---|---|---|
| | Probability: 0.999 AA1: | |
| 108662 | 25 AA2: 26 | MIIKPSITLLFFFALIISFPAKINA |
| | Probability: 1.000 AA1: | |
| 108671 | 23 AA2: 24 | MNSMNKKSVLLVFLLLLSSSLFA |
| | Probability: 0.519 AA1: | MSSCTPAGKSSSITLFLRFAFGNNAFAIRGAPLVSRKSP |
| 108673 | 72 AA2: 73 | GISTMVLLRQLMRRRLSLFTSAMITASRFSAAA |
| | Probability: 1.000 AA1: | MINLRMGHTYMMRFFKRRIARIKLVLAAVGIM |
| 108686 | 40 AA2: 41 | AAATQVWA |
| | Probability: 0.562 AA1: | |
| 108690 | 27 AA2: 28 | MILFPSAVISLSPCSSLYSFITLLTIA |

Any methods can be used for identifying "prepro" domain sequences and signal sequences, and many are well known in the art, see, e.g., Van de Ven (1993) Crit. Rev. Oncog. 4(2):115-136. For example, to identify a prepro sequence, the protein is purified from the extracellular space and the N-terminal protein sequence is determined and compared to the unprocessed form.

The invention includes polypeptides with or without a signal sequence and/or a prepro sequence. The invention includes polypeptides with heterologous signal sequences and/or prepro sequences. The prepro sequence (including a sequence of the invention used as a heterologous prepro domain) can be located on the amino terminal or the carboxy terminal end of the protein. The invention also includes isolated, synthetic or recombinant signal sequences, prepro sequences and catalytic domains (e.g., "active sites") comprising sequences of the invention. The polypeptide comprising a signal sequence of the invention can be a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention or another polypeptide, enzyme, protein, e.g. structural or binding protein, or another enzyme or other polypeptide.

The polypeptide, enzyme, protein, e.g. structural or binding protein signal sequences (SPs) and/or prepro sequences of the invention can be isolated peptides, or, sequences joined to another a polypeptide, enzyme, protein, e.g. structural or binding protein, or a non-polypeptide, non-enzyme, non-protein, e.g. non-structural or non-binding protein, e.g., as a fusion (chimeric) protein. In one aspect, the invention provides polypeptides comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, signal sequences of the invention. In one aspect, polypeptides comprising

polypeptide, enzyme, protein, e.g. structural or binding protein signal sequences SPs and/or prepro of the invention comprise sequences heterologous to a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention (e.g., a fusion protein comprising an SP and/or prepro of the invention and sequences from another a polypeptide, enzyme, protein, e.g. structural or binding protein, or a non-polypeptide, non-enzyme, non-protein, e.g. non-structural or non-binding protein). In one aspect, the invention provides a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention with heterologous SPs and/or prepro sequences, e.g., sequences with a yeast signal sequence. A polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can comprise a heterologous SP and/or prepro in a vector, e.g., a pPIC series vector (Invitrogen, Carlsbad, CA).

In one aspect, SPs and/or prepro sequences of the invention are identified following identification of novel a polypeptide, enzyme, protein, e.g. structural or binding protein. The pathways by which proteins are sorted and transported to their proper cellular location are often referred to as protein targeting pathways. One of the most important elements in all of these targeting systems is a short amino acid sequence at the amino terminus of a newly synthesized polypeptide called the signal sequence. This signal sequence directs a protein to its appropriate location in the cell and is removed during transport or when the protein reaches its final destination. Most lysosomal, membrane, or secreted proteins have an amino-terminal signal sequence that marks them for translocation into the lumen of the endoplasmic reticulum. The signal sequences can vary in length from about 10 to 65, or more, amino acid residues. Various methods of recognition of signal sequences are known to those of skill in the art, e.g., SignalP, as discussed above.

It should be understood that in some aspects a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention may not have SPs and/or prepro sequences, or "domains." In one aspect, the invention provides the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention lacking all or part of an SP and/or a prepro domain. In one aspect, the invention provides a nucleic acid sequence encoding a signal sequence (SP) and/or prepro from one a polypeptide, enzyme, protein, e.g. structural or binding protein, operably linked to a nucleic acid sequence of a different a polypeptide, enzyme, protein, e.g. structural or binding protein, or, optionally, a signal sequence (SPs) and/or prepro domain from a non-enzyme or non-protein, e.g. non-structural or non-binding protein, may be desired.

The invention also provides isolated, synthetic or recombinant polypeptides comprising signal sequences (SPs), prepro domain and/or catalytic domains (CDs) of the invention and heterologous sequences. The heterologous sequences are sequences not naturally associated (e.g., to a enzyme) with an SP, prepro domain and/or CD. The sequence to which the SP, prepro domain and/or CD are not naturally associated can be on the SP's, prepro domain and/or CD's amino terminal end, carboxy terminal end, and/or on both ends of the SP and/or CD. In one aspect, the invention provides an isolated, synthetic or recombinant polypeptide comprising (or consisting of) a polypeptide comprising a signal sequence (SP), prepro domain and/or catalytic domain (CD) of the invention with the proviso that it is not associated with any sequence to which it is naturally associated (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, sequence). Similarly in one aspect, the invention provides isolated, synthetic or recombinant nucleic acids encoding these polypeptides. Thus, in one aspect, the isolated, synthetic or recombinant nucleic acid of the invention comprises coding sequence for a signal sequence (SP), prepro domain and/or catalytic domain (CD) of the invention and a heterologous sequence (i.e., a sequence not naturally associated with the a signal sequence (SP), prepro domain and/or catalytic domain (CD) of the invention). The heterologous sequence can be on the 3' terminal end, 5' terminal end, and/or on both ends of the SP, prepro domain and/or CD coding sequence.

*Hybrid (chimeric) a polypeptide, enzyme, protein, e.g. structural or binding protein, and peptide libraries*

In one aspect, the invention provides hybrid a polypeptide, enzyme, protein, e.g. structural or binding protein, and fusion proteins, including peptide libraries, comprising sequences of the invention. The peptide libraries of the invention can be used to isolate peptide modulators (e.g., activators or inhibitors) of targets, such as a polypeptide, enzyme, protein, e.g. structural or binding protein, substrates, receptors, enzymes. The peptide libraries of the invention can be used to identify formal binding partners of targets, such as ligands, e.g., cytokines, hormones and the like. In one aspect, the invention provides chimeric proteins comprising a signal sequence (SP), prepro domain and/or catalytic domain (CD) of the invention or a combination thereof and a heterologous sequence (see above).

In one aspect, the fusion proteins of the invention (e.g., the peptide moiety) are conformationally stabilized (relative to linear peptides) to allow a higher binding affinity for targets. The invention provides fusions of a polypeptide, enzyme, protein, e.g.

structural or binding protein, of the invention and other peptides, including known and random peptides. They can be fused in such a manner that the structure of the polypeptide, enzyme, protein, e.g. structural or binding proteins is not significantly perturbed and the peptide is metabolically or structurally conformationally stabilized. This allows the creation of a peptide library that is easily monitored both for its presence within cells and its quantity.

Amino acid sequence variants of the invention can be characterized by a predetermined nature of the variation, a feature that sets them apart from a naturally occurring form, e.g., an allelic or interspecies variation of a polypeptide, enzyme, protein, e.g. structural or binding protein, sequence. In one aspect, the variants of the invention exhibit the same qualitative biological activity as the naturally occurring analogue. Alternatively, the variants can be selected for having modified characteristics. In one aspect, while the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed a polypeptide, enzyme, protein, e.g. structural or binding protein, variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, as discussed herein for example, M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants can be done using, e.g., assays of glucan hydrolysis. In alternative aspects, amino acid substitutions can be single residues; insertions can be on the order of from about 1 to 20 amino acids, although considerably larger insertions can be done. Deletions can range from about 1 to about 20, 30, 40, 50, 60, 70 residues or more. To obtain a final derivative with the optimal properties, substitutions, deletions, insertions or any combination thereof may be used. Generally, these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances.

The invention provides a polypeptide, enzyme, protein, e.g. structural or binding protein, where the structure of the polypeptide backbone, the secondary or the tertiary structure, e.g., an alpha-helical or beta-sheet structure, has been modified. In one aspect, the charge or hydrophobicity has been modified. In one aspect, the bulk of a side chain has been modified. Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative. For example, substitutions can be made which more significantly affect: the structure of the polypeptide backbone in the

area of the alteration, for example a alpha-helical or a beta-sheet structure; a charge or a hydrophobic site of the molecule, which can be at an active site; or a side chain. The invention provides substitutions in polypeptide of the invention where (a) a hydrophilic residues, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g. lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine. The variants can exhibit the same qualitative biological activity (i.e., an enzyme, structural or binding activity) although variants can be selected to modify the characteristics of the polypeptide, enzyme, protein, e.g. structural or binding proteins as needed.

In one aspect, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention comprise epitopes or purification tags, signal sequences or other fusion sequences, etc. In one aspect, the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention can be fused to a random peptide to form a fusion polypeptide. By "fused" or "operably linked" herein is meant that the random peptide and the polypeptide, enzyme, protein, e.g. structural or binding protein are linked together, in such a manner as to minimize the disruption to the stability of the polypeptide, enzyme, protein, e.g. structural or binding protein structure, e.g., it retains a polypeptide, enzyme, protein, e.g. structural or binding protein, activity. The fusion polypeptide (or fusion polynucleotide encoding the fusion polypeptide) can comprise further components as well, including multiple peptides at multiple loops.

In one aspect, the peptides and nucleic acids encoding them are randomized, either fully randomized or they are biased in their randomization, e.g. in nucleotide/residue frequency generally or per position. "Randomized" means that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. In one aspect, the nucleic acids which give rise to the peptides can be chemically synthesized, and thus may incorporate any nucleotide at any position. Thus, when the nucleic acids are expressed to form peptides, any amino acid residue may be incorporated at any position. The synthetic process can be designed to generate randomized nucleic acids, to allow the formation of all or most of the possible combinations over the length of the nucleic acid, thus forming a library of randomized nucleic acids. The library can provide a sufficiently structurally diverse population of randomized expression products to affect a

probabilistically sufficient range of cellular responses to provide one or more cells exhibiting a desired response. Thus, the invention provides an interaction library large enough so that at least one of its members will have a structure that gives it affinity for some molecule, protein, or other factor.

In one aspect, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention is a multidomain enzyme that comprises a signal peptide, a carbohydrate binding module, a polypeptide, enzyme, protein, e.g. structural or binding protein, catalytic domain, a linker and/or another catalytic domain.

The invention provides a means for generating chimeric polypeptides which may encode biologically active hybrid polypeptides (*e.g.*, hybrid a polypeptide, enzyme, protein, e.g. structural or binding protein, ). In one aspect, the original polynucleotides encode biologically active polypeptides. The method of the invention produces new hybrid polypeptides by utilizing cellular processes which integrate the sequence of the original polynucleotides such that the resulting hybrid polynucleotide encodes a polypeptide demonstrating activities derived from the original biologically active polypeptides. For example, the original polynucleotides may encode a particular enzyme from different microorganisms. An enzyme encoded by a first polynucleotide from one organism or variant may, for example, function effectively under a particular environmental condition, *e.g.* high salinity. An enzyme encoded by a second polynucleotide from a different organism or variant may function effectively under a different environmental condition, such as extremely high temperatures. A hybrid polynucleotide containing sequences from the first and second original polynucleotides may encode an enzyme which exhibits characteristics of both enzymes encoded by the original polynucleotides. Thus, the enzyme encoded by the hybrid polynucleotide may function effectively under environmental conditions shared by each of the enzymes encoded by the first and second polynucleotides, *e.g.*, high salinity and extreme temperatures.

A hybrid polypeptide resulting from the method of the invention may exhibit specialized enzyme activity, binding activity and/or structural activity not displayed in the original enzymes. For example, following recombination and/or reductive reassortment of polynucleotides encoding a polypeptide, enzyme, protein, e.g. structural or binding protein, the resulting hybrid polypeptide encoded by a hybrid polynucleotide can be screened for specialized non- enzyme, non-structural or non-binding activities, obtained from each of the original enzymes. Thus, for example, the hybrid polypeptide may be

screened to ascertain those chemical functionalities which distinguish the hybrid polypeptide from the original parent polypeptides, such as the temperature, pH or salt concentration at which the hybrid polypeptide functions.

In one aspect, the invention relates to a method for producing a biologically active hybrid polypeptide and screening such a polypeptide for enhanced activity by:

1) introducing at least a first polynucleotide in operable linkage and a second polynucleotide in operable linkage, the at least first polynucleotide and second polynucleotide sharing at least one region of partial sequence homology, into a suitable host cell;
2) growing the host cell under conditions which promote sequence reorganization resulting in a hybrid polynucleotide in operable linkage;
3) expressing a hybrid polypeptide encoded by the hybrid polynucleotide;
4) screening the hybrid polypeptide under conditions which promote identification of enhanced biological activity; and
5) isolating the a polynucleotide encoding the hybrid polypeptide.

Isolating, discovering polypeptides, enzymes, proteins, e.g. structural or binding protein

The invention provides methods for isolating and discovering a polypeptide, enzyme, protein, e.g. structural or binding protein, and the nucleic acids that encode them. Polynucleotides or enzymes may be isolated from individual organisms ("isolates"), collections of organisms that have been grown in defined media ("enrichment cultures"), or, uncultivated organisms ("environmental samples"). The organisms can be isolated by, e.g., *in vivo* biopanning (see discussion, below). The use of a culture-independent approach to derive polynucleotides encoding novel bioactivities from environmental samples is most preferable since it allows one to access untapped resources of biodiversity. Polynucleotides or enzymes also can be isolated from any one of numerous organisms, e.g. bacteria. In addition to whole cells, polynucleotides or enzymes also can be isolated from crude enzyme preparations derived from cultures of these organisms, e.g., bacteria.

"Environmental libraries" are generated from environmental samples and represent the collective genomes of naturally occurring organisms archived in cloning vectors that can be propagated in suitable prokaryotic hosts. Because the cloned DNA is initially extracted directly from environmental samples, the libraries are not limited to the small fraction of prokaryotes that can be grown in pure culture. Additionally, a normalization of the environmental DNA present in these samples could allow more

equal representation of the DNA from all of the species present in the original sample. This can dramatically increase the efficiency of finding interesting genes from minor constituents of the sample which may be under-represented by several orders of magnitude compared to the dominant species.

For example, gene libraries generated from one or more uncultivated microorganisms are screened for an activity of interest. Potential pathways encoding bioactive molecules of interest are first captured in prokaryotic cells in the form of gene expression libraries. Polynucleotides encoding activities of interest are isolated from such libraries and introduced into a host cell. The host cell is grown under conditions which promote recombination and/or reductive reassortment creating potentially active biomolecules with novel or enhanced activities.

*In vivo* biopanning may be performed utilizing a FACS-based and non-optical (e.g., magnetic) based machines. Complex gene libraries are constructed with vectors which contain elements which stabilize transcribed RNA. For example, the inclusion of sequences which result in secondary structures such as hairpins which are designed to flank the transcribed regions of the RNA would serve to enhance their stability, thus increasing their half life within the cell. The probe molecules used in the biopanning process consist of oligonucleotides labeled with reporter molecules that only fluoresce upon binding of the probe to a target molecule. These probes are introduced into the recombinant cells from the library using one of several transformation methods. The probe molecules bind to the transcribed target mRNA resulting in DNA/RNA heteroduplex molecules. Binding of the probe to a target will yield a fluorescent signal which is detected and sorted by the FACS machine during the screening process.

Additionally, subcloning may be performed to further isolate sequences of interest. In subcloning, a portion of DNA is amplified, digested, generally by restriction enzymes, to cut out the desired sequence, the desired sequence is ligated into a recipient vector and is amplified. At each step in subcloning, the portion is examined for the activity of interest, in order to ensure that DNA that encodes the structural protein has not been excluded. The insert may be purified at any step of the subcloning, for example, by gel electrophoresis prior to ligation into a vector or where cells containing the recipient vector and cells not containing the recipient vector are placed on selective media containing, for example, an antibiotic, which will kill the cells not containing the recipient vector. Specific methods of subcloning cDNA inserts into vectors are well-known in the art (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring

Harbor Laboratory Press (1989)). In another aspect, the enzymes of the invention are subclones. Such subclones may differ from the parent clone by, for example, length, a mutation, a tag or a label.

In one aspect, the signal sequences of the invention are identified following identification of a novel polypeptide, enzyme, protein, e.g. structural or binding protein. The pathways by which proteins are sorted and transported to their proper cellular location are often referred to as protein targeting pathways. One of the most important elements in all of these targeting systems is a short amino acid sequence at the amino terminus of a newly synthesized polypeptide called the signal sequence. This signal sequence directs a protein to its appropriate location in the cell and is removed during transport or when the protein reaches its final destination. Most lysosomal, membrane, or secreted proteins have an amino-terminal signal sequence that marks them for translocation into the lumen of the endoplasmic reticulum. More than 100 signal sequences for proteins in this group have been determined. The sequences vary in length from 13 to 36 amino acid residues. Various methods of recognition of signal sequences are known to those of skill in the art. In one aspect, the peptides are identified by a method referred to as SignalP. SignalP uses a combined neural network which recognizes both signal peptides and their cleavage sites. See, e.g., Nielsen (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering, vol. 10, no. 1, p. 1-6. It should be understood that some of the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention may or may not contain signal sequences. It may be desirable to include a nucleic acid sequence encoding a signal sequence from a polypeptide, enzyme, protein, e.g. structural or binding protein, operably linked to a nucleic acid sequence of a different a polypeptide, enzyme, protein, e.g. structural or binding protein may be desired.

The microorganisms from which the polynucleotide may be discovered, isolated or prepared include prokaryotic microorganisms, such as *Eubacteria* and *Archaebacteria* and lower eukaryotic microorganisms such as fungi, some algae and protozoa. Polynucleotides may be discovered, isolated or prepared from samples, such as environmental samples, in which case the nucleic acid may be recovered without culturing of an organism or recovered from one or more cultured organisms. In one aspect, such microorganisms may be extremophiles, such as hyperthermophiles, psychrophiles, psychrotrophs, halophiles, barophiles and acidophiles. Polynucleotides encoding enzymes isolated from extremophilic microorganisms can be used. Such

enzymes may function at temperatures above 100°C in terrestrial hot springs and deep searmal vents, at temperatures below 0°C in arctic waters, in the saturated salt environment of the Dead Sea, at pH values around 0 in coal deposits and geothermal sulfur-rich springs, or at pH values greater than 11 in sewage sludge. For example, several esterases and lipases cloned and expressed from extremophilic organisms show high activity throughout a wide range of temperatures and pHs.

Polynucleotides selected and isolated as hereinabove described are introduced into a suitable host cell. A suitable host cell is any cell which is capable of promoting recombination and/or reductive reassortment. The selected polynucleotides are in one aspect already in a vector which includes appropriate control sequences. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or in one aspect, the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis *et al.*, 1986).

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; and plant cells. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

With particular references to various mammalian cell culture systems that can be employed to express recombinant protein, examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981) and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

In another aspect, it is envisioned the method of the present invention can be used to generate novel polynucleotides encoding biochemical pathways from one or more

operons or gene clusters or portions thereof. For example, bacteria and many eukaryotes have a coordinated mechanism for regulating genes whose products are involved in related processes. The genes are clustered, in structures referred to as "gene clusters," on a single chromosome and are transcribed together under the control of a single regulatory sequence, including a single promoter which initiates transcription of the entire cluster. Thus, a gene cluster is a group of adjacent genes that are either identical or related, usually as to their function. An example of a biochemical pathway encoded by gene clusters are polyketides.

Gene cluster DNA can be isolated from different organisms and ligated into vectors, particularly vectors containing expression regulatory sequences which can control and regulate the production of a detectable protein or protein-related array activity from the ligated gene clusters. Use of vectors which have an exceptionally large capacity for exogenous DNA introduction are particularly appropriate for use with such gene clusters and are described by way of example herein to include the f-factor (or fertility factor) of *E. coli*. This f-factor of *E. coli* is a plasmid which affects high-frequency transfer of itself during conjugation and is ideal to achieve and stably propagate large DNA fragments, such as gene clusters from mixed microbial samples. One aspect is to use cloning vectors, referred to as "fosmids" or bacterial artificial chromosome (BAC) vectors. These are derived from *E. coli* f-factor which is able to stably integrate large segments of genomic DNA. When integrated with DNA from a mixed uncultured environmental sample, this makes it possible to achieve large genomic fragments in the form of a stable "environmental DNA library." Another type of vector for use in the present invention is a cosmid vector. Cosmid vectors were originally designed to clone and propagate large segments of genomic DNA. Cloning into cosmid vectors is described in detail in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press (1989). Once ligated into an appropriate vector, two or more vectors containing different polyketide synthase gene clusters can be introduced into a suitable host cell. Regions of partial sequence homology shared by the gene clusters will promote processes which result in sequence reorganization resulting in a hybrid gene cluster. The novel hybrid gene cluster can then be screened for enhanced activities not found in the original gene clusters.

Methods for screening for various enzyme activities are known to those of skill in the art and are discussed throughout the present specification. Such methods may be employed when isolating the polypeptides and polynucleotides of the invention.

In one aspect, the invention provides methods for discovering and isolating polypeptides, enzymes, proteins, e.g. structural or binding proteins or compounds to modify the enzymatic activity, using a whole cell approach. Putative clones encoding polypeptides, enzymes, proteins, e.g. structural or binding proteins from genomic DNA library can be screened.

Screening Methodologies and "On-line" Monitoring Devices

In practicing the methods of the invention, a variety of apparatus and methodologies can be used to in conjunction with the polypeptides and nucleic acids of the invention, e.g., to screen polypeptides for a polypeptide, enzyme, protein, e.g. structural or binding protein, activity, to screen compounds as potential modulators, e.g., activators or inhibitors, of an enzyme, structural or binding activity, for antibodies that bind to a polypeptide of the invention, for nucleic acids that hybridize to a nucleic acid of the invention, to screen for cells expressing a polypeptide of the invention and the like. In addition to the array formats described in detail below for screening samples, alternative formats can also be used to practice the methods of the invention. Such formats include, for example, mass spectrometers, chromatographs, e.g., high-throughput HPLC and other forms of liquid chromatography, and smaller formats, such as 1536-well plates, 384–well plates and so on. High throughput screening apparatus can be adapted and used to practice the methods of the invention, see, e.g., U.S. Patent Application No. 20020001809.

*Capillary Arrays*

Nucleic acids or polypeptides of the invention can be immobilized to or applied to an array. Arrays can be used to screen for or monitor libraries of compositions (e.g., small molecules, antibodies, nucleic acids, etc.) for their ability to bind to or modulate the activity of a nucleic acid or a polypeptide of the invention. Capillary arrays, such as the GIGAMATRIX™, Diversa Corporation, San Diego, CA; and arrays described in, e.g., U.S. Patent Application No. 20020080350 A1; WO 0231203 A; WO 0244336 A, provide an alternative apparatus for holding and screening samples. In one aspect, the capillary array includes a plurality of capillaries formed into an array of adjacent capillaries, wherein each capillary comprises at least one wall defining a lumen for retaining a sample. The lumen may be cylindrical, square, hexagonal or any other geometric shape so long as the walls form a lumen for retention of a liquid or sample. The capillaries of the capillary array can be held together in close proximity to form a planar structure. The

capillaries can be bound together, by being fused (e.g., where the capillaries are made of glass), glued, bonded, or clamped side-by-side. Additionally, the capillary array can include interstitial material disposed between adjacent capillaries in the array, thereby forming a solid planar device containing a plurality of through-holes.

A capillary array can be formed of any number of individual capillaries, for example, a range from 100 to 4,000,000 capillaries. Further, a capillary array having about 100,000 or more individual capillaries can be formed into the standard size and shape of a Microtiter® plate for fitment into standard laboratory equipment. The lumens are filled manually or automatically using either capillary action or microinjection using a thin needle. Samples of interest may subsequently be removed from individual capillaries for further analysis or characterization. For example, a thin, needle-like probe is positioned in fluid communication with a selected capillary to either add or withdraw material from the lumen.

In a single-pot screening assay, the assay components are mixed yielding a solution of interest, prior to insertion into the capillary array. The lumen is filled by capillary action when at least a portion of the array is immersed into a solution of interest. Chemical or biological reactions and/or activity in each capillary are monitored for detectable events. A detectable event is often referred to as a "hit", which can usually be distinguished from "non-hit" producing capillaries by optical detection. Thus, capillary arrays allow for massively parallel detection of "hits".

In a multi-pot screening assay, a polypeptide or nucleic acid, e.g., a ligand, can be introduced into a first component, which is introduced into at least a portion of a capillary of a capillary array. An air bubble can then be introduced into the capillary behind the first component. A second component can then be introduced into the capillary, wherein the second component is separated from the first component by the air bubble. The first and second components can then be mixed by applying hydrostatic pressure to both sides of the capillary array to collapse the bubble. The capillary array is then monitored for a detectable event resulting from reaction or non-reaction of the two components.

In a binding screening assay, a sample of interest can be introduced as a first liquid labeled with a detectable particle into a capillary of a capillary array, wherein the lumen of the capillary is coated with a binding material for binding the detectable particle to the lumen. The first liquid may then be removed from the capillary tube, wherein the bound detectable particle is maintained within the capillary, and a second liquid may be

introduced into the capillary tube. The capillary is then monitored for a detectable event resulting from reaction or non-reaction of the particle with the second liquid.

*Arrays, or "Biochips"*

Nucleic acids or polypeptides of the invention can be immobilized to or applied to an array. Arrays can be used to screen for or monitor libraries of compositions (e.g., small molecules, antibodies, nucleic acids, etc.) for their ability to bind to or modulate the activity of a nucleic acid or a polypeptide of the invention. For example, in one aspect of the invention, a monitored parameter is transcript expression of a polypeptide, enzyme, protein, e.g. structural or binding protein, gene. One or more, or, all the transcripts of a cell can be measured by hybridization of a sample comprising transcripts of the cell, or, nucleic acids representative of or complementary to transcripts of a cell, by hybridization to immobilized nucleic acids on an array, or "biochip." By using an "array" of nucleic acids on a microchip, some or all of the transcripts of a cell can be simultaneously quantified. Alternatively, arrays comprising genomic nucleic acid can also be used to determine the genotype of a newly engineered strain made by the methods of the invention. Polypeptide arrays" can also be used to simultaneously quantify a plurality of proteins. The present invention can be practiced with any known "array," also referred to as a "microarray" or "nucleic acid array" or "polypeptide array" or "antibody array" or "biochip," or variation thereof. Arrays are generically a plurality of "spots" or "target elements," each target element comprising a defined amount of one or more biological molecules, e.g., oligonucleotides, immobilized onto a defined area of a substrate surface for specific binding to a sample molecule, e.g., mRNA transcripts.

The terms "array" or "microarray" or "biochip" or "chip" as used herein is a plurality of target elements, each target element comprising a defined amount of one or more polypeptides (including antibodies) or nucleic acids immobilized onto a defined area of a substrate surface, as discussed in further detail, below.

In practicing the methods of the invention, any known array and/or method of making and using arrays can be incorporated in whole or in part, or variations thereof, as described, for example, in U.S. Patent Nos. 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; 5,434,049; see also, e.g., WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; see also, e.g., Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124;

Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. See also published U.S. patent applications Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

Antibodies and Antibody-based screening methods

The invention provides isolated, synthetic or recombinant antibodies that specifically bind to a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention. These antibodies can be used to isolate, identify or quantify the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention or related polypeptides. These antibodies can be used to isolate other polypeptides within the scope the invention or other related a polypeptide, enzyme, protein, e.g. structural or binding protein, . The antibodies can be designed to bind to an active site of a polypeptide, enzyme, protein, e.g. structural or binding protein. Thus, the invention provides methods of inhibiting a polypeptide, enzyme, protein, e.g. structural or binding protein, using the antibodies of the invention (see discussion above regarding applications for anti-polypeptide, anti-enzyme, anti-protein, e.g., anti-structural or anti-binding protein compositions of the invention).

The invention provides fragments of the enzymes of the invention, including immunogenic fragments of a polypeptide of the invention. The invention provides compositions comprising a polypeptide or peptide of the invention and adjuvants or carriers and the like.

The term "antibody" includes a peptide or polypeptide derived from, modeled after or substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, capable of specifically binding an antigen or epitope, see, e.g. Fundamental Immunology, Third Edition, W.E. Paul, ed., Raven Press, N.Y. (1993); Wilson (1994) J. Immunol. Methods 175:267-273; Yarmush (1992) J. Biochem. Biophys. Methods 25:85-97. The term antibody includes antigen-binding portions, i.e., "antigen binding sites," (e.g., fragments, subsequences, complementarity determining regions (CDRs)) that retain capacity to bind antigen, including (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a F(ab')2 fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward et al., (1989) Nature 341:544-546), which consists of a VH domain; and

(vi) an isolated complementarity determining region (CDR). Single chain antibodies are also included by reference in the term "antibody."

The antibodies can be used in immunoprecipitation, staining, immunoaffinity columns, and the like. If desired, nucleic acid sequences encoding for specific antigens can be generated by immunization followed by isolation of polypeptide or nucleic acid, amplification or cloning and immobilization of polypeptide onto an array of the invention. Alternatively, the methods of the invention can be used to modify the structure of an antibody produced by a cell to be modified, e.g., an antibody's affinity can be increased or decreased. Furthermore, the ability to make or modify antibodies can be a phenotype engineered into a cell by the methods of the invention.

Methods of immunization, producing and isolating antibodies (polyclonal and monoclonal) are known to those of skill in the art and described in the scientific and patent literature, see, e.g., Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2d ed.) Academic Press, New York, NY (1986); Kohler (1975) Nature 256:495; Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York. Antibodies also can be generated *in vitro*, e.g., using recombinant antibody binding site expressing phage display libraries, in addition to the traditional in vivo methods using animals. See, e.g., Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45.

The polypeptides of the invention or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may also be used to generate antibodies which bind specifically to the polypeptides or fragments. The resulting antibodies may be used in immunoaffinity chromatography procedures to isolate or purify the polypeptide or to determine whether the polypeptide is present in a biological sample. In such procedures, a protein preparation, such as an extract, or a biological sample is contacted with an antibody capable of specifically binding to one of the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof.

In immunoaffinity procedures, the antibody is attached to a solid support, such as a bead or other column matrix. The protein preparation is placed in contact with the antibody under conditions in which the antibody specifically binds to one of the

polypeptides of the invention, or fragment thereof. After a wash to remove non-specifically bound proteins, the specifically bound polypeptides are eluted.

The ability of proteins in a biological sample to bind to the antibody may be determined using any of a variety of procedures familiar to those skilled in the art. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. Particular assays include ELISA assays, sandwich assays, radioimmunoassays and Western Blots.

Polyclonal antibodies generated against the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, for example, a nonhuman. The antibody so obtained can bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies which may bind to the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from cells expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, Nature, 256:495-497, 1975), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today 4:72, 1983) and the EBV-hybridoma technique (Cole, *et al.*, 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. Alternatively, transgenic mice may be used to express humanized antibodies to these polypeptides or fragments thereof.

Antibodies generated against the polypeptides of the invention, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof may be used in screening for similar polypeptides from other organisms and samples. In such techniques, polypeptides from the organism are contacted with the antibody and those polypeptides which specifically bind the antibody are detected. Any of the procedures described above may be used to detect antibody binding. One such

screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in Enzymology*, Vol 160, pp. 87-116.

Kits

The invention provides kits comprising the compositions, e.g., nucleic acids, expression cassettes, vectors, cells, transgenic seeds or plants or plant parts, polypeptides and/or antibodies of the invention. The kits also can contain instructional material teaching the methodologies and uses of the enzymes of the invention (see, e.g., Table 1, Table 3 and/or Table 4), including the industrial, experimental, food and feed processing and medical uses of the compositions and methods of the invention, as described herein.

Whole cell engineering and measuring metabolic parameters

The methods of the invention provide whole cell evolution, or whole cell engineering, of a cell to develop a new cell strain having a new phenotype, e.g., a new or modified a polypeptide, enzyme, protein, e.g. structural or binding protein, activity, by modifying the genetic composition of the cell. The genetic composition can be modified by addition to the cell of a nucleic acid of the invention, e.g., a coding sequence for an enzyme of the invention. See, e.g., WO0229032; WO0196551.

To detect the new phenotype, at least one metabolic parameter of a modified cell is monitored in the cell in a "real time" or "on-line" time frame. In one aspect, a plurality of cells, such as a cell culture, is monitored in "real time" or "on-line." In one aspect, a plurality of metabolic parameters is monitored in "real time" or "on-line." Metabolic parameters can be monitored using the polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention.

Metabolic flux analysis (MFA) is based on a known biochemistry framework. A linearly independent metabolic matrix is constructed based on the law of mass conservation and on the pseudo-steady state hypothesis (PSSH) on the intracellular metabolites. In practicing the methods of the invention, metabolic networks are established, including the:

- identity of all pathway substrates, products and intermediary metabolites
- identity of all the chemical reactions interconverting the pathway metabolites, the stoichiometry of the pathway reactions,
- identity of all the enzymes catalyzing the reactions, the enzyme reaction kinetics,
- the regulatory interactions between pathway components, e.g. allosteric interactions, enzyme-enzyme interactions etc,

• intracellular compartmentalization of enzymes or any other supramolecular organization of the enzymes, and,

• the presence of any concentration gradients of metabolites, enzymes or effector molecules or diffusion barriers to their movement.

Once the metabolic network for a given strain is built, mathematic presentation by matrix notion can be introduced to estimate the intracellular metabolic fluxes if the on-line metabolome data is available. Metabolic phenotype relies on the changes of the whole metabolic network within a cell. Metabolic phenotype relies on the change of pathway utilization with respect to environmental conditions, genetic regulation, developmental state and the genotype, etc. In one aspect of the methods of the invention, after the on-line MFA calculation, the dynamic behavior of the cells, their phenotype and other properties are analyzed by investigating the pathway utilization. For example, if the glucose supply is increased and the oxygen decreased during the yeast fermentation, the utilization of respiratory pathways will be reduced and/or stopped, and the utilization of the fermentative pathways will dominate. Control of physiological state of cell cultures will become possible after the pathway analysis. The methods of the invention can help determine how to manipulate the fermentation by determining how to change the substrate supply, temperature, use of inducers, etc. to control the physiological state of cells to move along desirable direction. In practicing the methods of the invention, the MFA results can also be compared with transcriptome and proteome data to design experiments and protocols for metabolic engineering or gene shuffling, etc.

In practicing the methods of the invention, any modified or new phenotype can be conferred and detected, including new or improved characteristics in the cell. Any aspect of metabolism or growth can be monitored.

*Monitoring expression of an mRNA transcript*

In one aspect of the invention, the engineered phenotype comprises increasing or decreasing the expression of an mRNA transcript (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein, message) or generating new (e.g., polypeptide, enzyme, protein, e.g. structural or binding protein) transcripts in a cell. This increased or decreased expression can be traced by testing for the presence of a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention or by a polypeptide, enzyme, protein, e.g. structural or binding protein, activity assays. Such assays are well known in the art. mRNA transcripts, or messages, also can be detected and quantified by any method known in the art, including, e.g., Northern blots, quantitative amplification

reactions, hybridization to arrays, and the like. Quantitative amplification reactions include, e.g., quantitative PCR, including, e.g., quantitative reverse transcription polymerase chain reaction, or RT-PCR; quantitative real time RT-PCR, or "real-time kinetic RT-PCR" (see, e.g., Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

In one aspect of the invention, the engineered phenotype is generated by knocking out expression of a homologous gene. The gene's coding sequence or one or more transcriptional control elements can be knocked out, e.g., promoters or enhancers. Thus, the expression of a transcript can be completely ablated or only decreased.

In one aspect of the invention, the engineered phenotype comprises increasing the expression of a homologous gene. This can be effected by knocking out of a negative control element, including a transcriptional regulatory element acting in cis- or trans- , or, mutagenizing a positive control element. One or more, or, all the transcripts of a cell can be measured by hybridization of a sample comprising transcripts of the cell, or, nucleic acids representative of or complementary to transcripts of a cell, by hybridization to immobilized nucleic acids on an array.

*Monitoring expression of a polypeptides, peptides and amino acids*

In one aspect of the invention, the engineered phenotype comprises increasing or decreasing the expression of a polypeptide (e.g., a polypeptide, enzyme, protein, e.g. structural or binding protein,) or generating new polypeptides in a cell. This increased or decreased expression can be traced by determining the amount of a polypeptide, enzyme, protein, e.g. structural or binding protein, present or by a polypeptide, enzyme, protein, e.g. structural or binding protein, activity assays.

Polypeptides, peptides and amino acids also can be detected and quantified by any method known in the art, including, e.g., nuclear magnetic resonance (NMR), spectrophotometry, radiography (protein radiolabeling), electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, various immunological methods, e.g. immunoprecipitation, immunodiffusion, immuno-electrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, gel electrophoresis (e.g., SDS-PAGE), staining with antibodies, fluorescent activated cell sorter (FACS), pyrolysis mass spectrometry, Fourier-Transform Infrared Spectrometry, Raman spectrometry, GC-MS, and LC-Electrospray and cap-LC-tandem-electrospray mass spectrometries, and the like. Novel bioactivities can also be

screened using methods, or variations thereof, described in U.S. Patent No. 6,057,103. Furthermore, as discussed below in detail, one or more, or, all the polypeptides of a cell can be measured using a protein array.

Pharmaceutical Compositions and Dietary Supplements

The invention provides pharmaceutical compositions, e.g., formulations, comprising a composition (including polypeptide, nucleic acid, or antibody) of the invention and a pharmaceutically acceptable excipient. The invention provides enteral and parenteral formulations comprising compositions of the invention. For example, the invention provides oral formulations (including or dietary supplements) comprising a composition of the invention. The invention provides formulations and methods for treating, ameliorating, diagnosing or preventing disease of condition of interest; e.g., in one aspect the invention provides methods comprising providing a pharmaceutical composition or dietary supplement comprising a composition of the invention; and administering an effective amount of the pharmaceutical composition or dietary supplement to a subject in need thereof.

The compositions and methods of the invention can also be practiced *ex vivo* or *in vitro*, or on a non-biological fluid or substance. In one aspect, the compositions and methods comprise providing a pharmaceutical composition or dietary supplement comprising a formulation of the invention; and administering an effective amount of the pharmaceutical composition or dietary supplement to a subject in need thereof.

The pharmaceutical compositions and dietary supplements used in the methods of the invention can be administered by any means known in the art, e.g., parenterally, topically, orally, or by local administration, such as by aerosol or transdermally. The compositions and dietary supplements of the invention can be formulated as a tablet, gel, geltab, pill, implant, liquid, spray, powder, food, feed pellet, as an injectable formulation or as an encapsulated formulation. The pharmaceutical compositions and dietary supplements can be formulated in any way and can be administered in a variety of unit dosage forms depending upon the condition or disease and the degree of illness, the general medical condition of each patient, the resulting preferred method of administration and the like. Details on techniques for formulation and administration are well described in the scientific and patent literature, see, e.g., the latest edition of Remington's Pharmaceutical Sciences, Maack Publishing Co, Easton PA ("Remington's") (e.g., Remington, The Science and Practice of Pharmacy, 21st Edition, by University of the Sciences in Philadelphia, Editor).

Pharmaceutical formulations and dietary supplements can be prepared according to any method known to the art for the manufacture of pharmaceuticals and dietary supplements. Such drugs and dietary supplements can contain sweetening agents, flavoring agents, coloring agents and preserving agents. A formulation (which includes "dietary supplements") can be admixtured with nontoxic pharmaceutically or orally acceptable excipients which are suitable for manufacture. Formulations may comprise one or more diluents, emulsifiers, preservatives, buffers, excipients, etc. and may be provided in such forms as liquids, powders, emulsions, lyophilized powders, sprays, creams, lotions, controlled release formulations, tablets, pills, gels, on patches, in implants, etc.

Pharmaceutical formulations and dietary supplements for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in appropriate and suitable dosages. Such carriers enable the pharmaceuticals and dietary supplements to be formulated in unit dosage forms as tablets, pills, powder, dragees, capsules, liquids, lozenges, gels, syrups, slurries, suspensions, etc., suitable for ingestion by the patient. Pharmaceutical preparations and dietary supplements for oral use can be formulated as a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable additional compounds, if desired, to obtain tablets or dragee cores. Suitable solid excipients are carbohydrate or protein fillers include, e.g., sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxy-methylcellulose; and gums including arabic and tragacanth; and proteins, e.g., gelatin and collagen. Disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound (i.e., dosage). Pharmaceutical preparations and dietary supplements of the invention can also be used orally using, e.g., push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active agents mixed with a filler or binders such as lactose or starches, lubricants

such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active agents can be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Aqueous suspensions can contain an active agent (e.g., an enzyme or peptidomimetic of the invention) in admixture with excipients suitable for the manufacture of aqueous suspensions. Such excipients include a suspending agent, such as sodium carboxymethylcellulose, methylcellulose, hydroxypropylmethylcellulose, sodium alginate, polyvinylpyrrolidone, gum tragacanth and gum acacia, and dispersing or wetting agents such as a naturally occurring phosphatide (e.g., lecithin), a condensation product of an alkylene oxide with a fatty acid (e.g., polyoxyethylene stearate), a condensation product of ethylene oxide with a long chain aliphatic alcohol (e.g., heptadecaethylene oxycetanol), a condensation product of ethylene oxide with a partial ester derived from a fatty acid and a hexitol (e.g., polyoxyethylene sorbitol mono-oleate), or a condensation product of ethylene oxide with a partial ester derived from fatty acid and a hexitol anhydride (e.g., polyoxyethylene sorbitan mono-oleate). The aqueous suspension can also contain one or more preservatives such as ethyl or n-propyl p-hydroxybenzoate, one or more coloring agents, one or more flavoring agents and one or more sweetening agents, such as sucrose, aspartame or saccharin. Formulations can be adjusted for osmolarity.

Oil-based pharmaceuticals are particularly useful for administration of hydrophobic formulations or active agents of the invention. Oil-based suspensions can be formulated by suspending an active agent (e.g., a composition of the invention) in a vegetable oil, such as arachis oil, olive oil, sesame oil or coconut oil, or in a mineral oil such as liquid paraffin; or a mixture of these. See e.g., U.S. Patent No. 5,716,928 describing using essential oils or essential oil components for increasing bioavailability and reducing inter- and intra-individual variability of orally administered hydrophobic pharmaceutical compounds (see also U.S. Patent No. 5,858,401). The oil suspensions can contain a thickening agent, such as beeswax, hard paraffin or cetyl alcohol. Sweetening agents can be added to provide a palatable oral preparation, such as glycerol, sorbitol or sucrose. These formulations and dietary supplements can be preserved by the addition of an antioxidant such as ascorbic acid. As an example of an injectable oil vehicle, see Minto (1997) J. Pharmacol. Exp. Ther. 281:93-102.

The pharmaceutical formulations and dietary supplements of the invention can also be in the form of oil-in-water emulsions. The oily phase can be a vegetable oil or a

mineral oil, described above, or a mixture of these. Suitable emulsifying agents include naturally-occurring gums, such as gum acacia and gum tragacanth, naturally occurring phosphatides, such as soybean lecithin, esters or partial esters derived from fatty acids and hexitol anhydrides, such as sorbitan mono-oleate, and condensation products of these partial esters with ethylene oxide, such as polyoxyethylene sorbitan mono-oleate. The emulsion can also contain sweetening agents and flavoring agents, as in the formulation of syrups and elixirs. Such formulations can also contain a demulcent, a preservative, or a coloring agent.

In the methods of the invention, the pharmaceutical compounds and dietary supplements can also be administered by in intranasal, intraocular and intravaginal routes including suppositories, insufflation, powders and aerosol formulations (for examples of steroid inhalants, see Rohatagi (1995) J. Clin. Pharmacol. 35:1187-1193; Tjwa (1995) Ann. Allergy Asthma Immunol. 75:107-111). Suppositories formulations can be prepared by mixing the drug with a suitable non-irritating excipient which is solid at ordinary temperatures but liquid at body temperatures and will therefore melt in the body to release the drug. Such materials are cocoa butter and polyethylene glycols.

In the methods of the invention, the pharmaceutical compounds and dietary supplements can be delivered by transdermally, by a topical route, formulated as applicator sticks, solutions, suspensions, emulsions, gels, creams, ointments, pastes, jellies, paints, powders, and aerosols.

In the methods of the invention, the pharmaceutical compounds and dietary supplements can also be delivered as microspheres for slow release in the body. For example, microspheres can be administered via intradermal injection of drug which slowly release subcutaneously; see Rao (1995) J. Biomater Sci. Polym. Ed. 7:623-645; as biodegradable and injectable gel formulations, see, e.g., Gao (1995) Pharm. Res. 12:857-863 (1995); or, as microspheres for oral administration, see, e.g., Eyles (1997) J. Pharm. Pharmacol. 49:669-674.

In the methods of the invention, the pharmaceutical compounds can be parenterally administered, such as by intravenous (IV) administration or administration into a body cavity or lumen of an organ. These formulations can comprise a solution of active agent dissolved in a pharmaceutically acceptable carrier. Acceptable vehicles and solvents that can be employed are water and Ringer's solution, an isotonic sodium chloride. In addition, sterile fixed oils can be employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono-

or diglycerides. In addition, fatty acids such as oleic acid can likewise be used in the preparation of injectables. These solutions are sterile and generally free of undesirable matter. These formulations may be sterilized by conventional, well known sterilization techniques. The formulations may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents, e.g., sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate and the like. The concentration of active agent in these formulations can vary widely, and will be selected primarily based on fluid volumes, viscosities, body weight, and the like, in accordance with the particular mode of administration selected and the patient's needs. For IV administration, the formulation can be a sterile injectable preparation, such as a sterile injectable aqueous or oleaginous suspension. This suspension can be formulated using those suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation can also be a suspension in a nontoxic parenterally-acceptable diluent or solvent, such as a solution of 1,3-butanediol. The administration can be by bolus or continuous infusion (e.g., substantially uninterrupted introduction into a blood vessel for a specified period of time).

The pharmaceutical compounds, formulations and dietary supplements of the invention can be lyophilized. The invention provides a stable lyophilized formulation comprising a composition of the invention, which can be made by lyophilizing a solution comprising a pharmaceutical of the invention and a bulking agent, e.g., mannitol, trehalose, raffinose, and sucrose or mixtures thereof. A process for preparing a stable lyophilized formulation can include the equivalent of lyophilizing a solution about 2.5 mg/mL protein, about 15 mg/mL sucrose, about 19 mg/mL NaCl, and a sodium citrate buffer having a pH greater than 5.5 but less than 6.5. See, e.g., U.S. patent app. no. 20040028670.

The compositions (e.g., formulations, including dietary supplements) of the invention can be delivered by the use of liposomes. By using liposomes, particularly where the liposome surface carries ligands specific for target cells, or are otherwise preferentially directed to a specific organ, one can focus the delivery of the active agent into target cells *in vivo*. See, e.g., U.S. Patent Nos. 6,063,400; 6,007,839; Al-Muhammed (1996) J. Microencapsul. 13:293-306; Chonn (1995) Curr. Opin. Biotechnol. 6:698-708; Ostro (1989) Am. J. Hosp. Pharm. 46:1576-1587.

The compositions (e.g., formulations, including dietary supplements) of the invention can be administered for prophylactic and/or therapeutic treatments. In

therapeutic applications, compositions are administered to a subject already suffering from a condition, infection or disease of interest in an amount sufficient to cure, alleviate or partially arrest the clinical manifestations of the condition, infection or disease and its complications (a "therapeutically effective amount"). In the methods of the invention, a pharmaceutical composition is administered in an amount sufficient to treat (e.g., ameliorate) or prevent a disease-related condition, a diseases or a symptoms, or to decrease or increase the amount of substance in a body fluid such as blood, serum, CSF and the like. The amount of composition (e.g., pharmaceutical compositions, formulations, including dietary supplements) adequate to accomplish this is defined as a "therapeutically effective dose." The dosage schedule and amounts effective for this use, i.e., the "dosing regimen," will depend upon a variety of factors, including the stage of the disease or condition, the severity of the disease or condition, the general state of the patient's health, the patient's physical status, age and the like. In calculating the dosage regimen for a patient, the mode of administration also is taken into consideration.

The dosage regimen also takes into consideration pharmacokinetics parameters well known in the art, i.e., the active agents' rate of absorption, bioavailability, metabolism, clearance, and the like (see, e.g., Hidalgo-Aragones (1996) J. Steroid Biochem. Mol. Biol. 58:611-617; Groning (1996) Pharmazie 51:337-341; Fotherby (1996) Contraception 54:59-69; Johnson (1995) J. Pharm. Sci. 84:1144-1146; Rohatagi (1995) Pharmazie 50:610-613; Brophy (1983) Eur. J. Clin. Pharmacol. 24:103-108; the latest Remington's, supra). The state of the art allows the clinician to determine the dosage regimen for each individual patient, active agent and disease or condition treated. Guidelines provided for similar compositions used as pharmaceuticals can be used as guidance to determine the dosage regiment, i.e., dose schedule and dosage levels, administered practicing the methods of the invention are correct and appropriate.

Single or multiple administrations of compositions (e.g., pharmaceutical compositions, formulations, including dietary supplements) of the invention can be given depending on the dosage and frequency as required and tolerated by the patient. The compositions should provide a sufficient quantity of active agent to effectively treat, ameliorate or prevent PKU or other PKU-related conditions, diseases or symptoms. For example, an exemplary pharmaceutical formulation for oral administration of a protein of the invention is in a daily amount of between about 0.1 to 0.5 to about 20, 50, 100 or 1000 or more *u*g per kilogram of body weight per day. In an alternative embodiment, dosages are from about 1 mg to about 4 mg per kg of body weight per patient per day are

used. Lower dosages can be used, in contrast to administration orally, into the blood stream, into a body cavity or into a lumen of an organ. Substantially higher dosages can be used in topical or oral administration or administering by powders, spray or inhalation. Actual methods for preparing parenterally or non-parenterally administrable formulations will be known or apparent to those skilled in the art and are described in more detail in such publications as Remington's, supra.

The compositions (e.g., pharmaceutical compositions, formulations, including dietary supplements) of the invention can further comprise other drugs or pharmaceuticals, e.g., compositions for treating a disease of interest and related symptoms or conditions. The methods of the invention can further comprise co-administration with other drugs or pharmaceuticals, e.g., compositions for treating septic shock, infection, fever, pain and related symptoms or conditions. For example, the methods and/or compositions and formulations of the invention can be co-formulated with and/or co-administered with antibiotics (e.g., antibacterial or bacteriostatic peptides or proteins), particularly those effective against bacteria or toxins, e.g., germ warfare agents, gram negative bacteria, fluids, cytokines, immunoregulatory agents, anti-inflammatory agents, complement activating agents, such as peptides or proteins comprising collagen-like domains or fibrinogen-like domains (e.g., a ficolin), carbohydrate-binding domains, and the like and combinations thereof.

In one aspect, the polypeptide (e.g., including a pharmaceutical composition or dietary supplement) of the invention is chemically modified. For example, the polypeptide can be chemically modified to produce a protected form that possesses better specific activity, prolonged half-life, and/or reduced immunogenicity *in vivo*. A polypeptide of the invention can be modified by any means known in the art, for example, by glycosylation, pegylation or a combination thereof.

In one aspect, the polypeptide (e.g., including a pharmaceutical composition or dietary supplement) of the invention is formulated by encapsulation in a liposome, or a micro- or nano-structure, such as a nanotubule or a nano- or microcapsule.

In one aspect, the polypeptide is formulated in a matrix stabilized enzyme crystal. The invention also provides matrix stabilized enzyme crystals comprising a polypeptide of the invention for use as pharmaceutical composition or dietary supplement, e.g., to treat or ameliorate a disease or condition of interest, e.g., as described in U.S. Patent App. No. 20020182201; for example, the formulation can be a cross-linked crystalline enzyme and a polymer with a reactive moiety effective to adhere to the crystal layer of the

crystalline enzyme. The invention also provides polypeptides of the invention as polymers in the form of multimerized (e.g., multi-functional) cross-linking forms; which in one aspect comprise a matrix stabilized enzyme crystal, e.g., a form resistant to degradation by proteolytic enzymes; and in alternative aspects, the cross-linking reagents comprise a dialdehyde cross-linking reagent, such as a linear or branched dialdehyde, or a substituted or unsubstituted glutaraldehyde (1,5-pentanedial), malonaldehyde (1,3-propanedial), succinaldehyde (1,4-butanedial), adipaldehyde (1,6-hexanedial), pimelaldehyde (1,7-heptanedial), or, glutaraldehyde; in other alternative aspects, the cross-linking reagents comprise carbodiimides, isoxazolium derivatives, chloroformates, carbonyldiimidazole, bis-imidoesters, bis-succinimidyl derivatives, di-isocyanates, di-isothiocyanates, di-sylfonyl halides, bis-nitrophenyl esters, dialdehydes, diacylazides, bis-maleimides, bis-haloacetyl derivatives, di-alkyl halides and bis-oxiranes (e.g., as described in U.S. Pat. No. 5,753,487).

The compositions of the invention can also be manufactured into biocompatible matrices, e.g., sol-gels, for encapsulating a polypeptide of the invention for use as pharmaceutical composition or dietary supplement, e.g., to treat or ameliorate a disease or condition of interest. In one aspect, compositions of the invention are manufactured as silica-based (e.g., oxysilane) sol-gel matrices, e.g., as described in U.S. Pat. No. 6,395,299, Pat. App. No. 20040241205. The invention also provides nano- or microcapsules comprising a composition of the invention for use as pharmaceutical composition or dietary supplement, e.g., to treat or ameliorate a disease or condition of interest, e.g., as described in U.S. Patent App. No. 20030157181.

The pharmaceutical compositions of the invention can be manufactured using any conventional method, e.g., mixing, dissolving, granulating, dragée-making, levigating, emulsifying, encapsulating, entrapping, melt-spinning, spray-drying, or lyophilizing processes. Alternative pharmaceutical formulations can be determined depending on the patient (e.g., adult or pediatric), condition, route of administration (e.g., oral) and the desired dosage.

Applications – Industrial, Medical, Experimental, Food and Feed Processing

Polypeptides (including enzymes and antibodies) and nucleic acids of the invention can be used for a variety of industrial, experimental, food and feed processing, nutritional and pharmaceutical applications, e.g., for food and feed supplements, colorants, neutraceuticals, cosmetic and pharmaceutical needs, making fuels such as bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel and the like.

Polypeptides of the invention (e.g., having enzyme, structural or binding activity) can be highly selective catalysts. The invention provides methods using enzymes of the invention in the food and feed industries, e.g., in methods for making food and feed products and food and feed additives. In one aspect, the invention provides processes using enzymes of the invention in the medical industry, e.g., to make pharmaceuticals.

The enzymes of the invention can catalyze reactions with exquisite stereo-, regio- and chemo- selectivities. The polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention can be engineered to function in various solvents, operate at extreme pHs (for example, high pHs and low pHs) extreme temperatures (for example, high temperatures and low temperatures), extreme salinity levels (for example, high salinity and low salinity) and catalyze reactions with compounds that are structurally unrelated to their natural, physiological substrates.

*Animal feeds and food or feed additives*

The invention provides compositions (e.g., enzymes of the invention, as those described in Tables 1, 2, and 3) methods for treating animal feeds and foods and food or feed additives using a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention, and/or the antibodies of the invention. The invention provides animal feeds, foods, and additives comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention, antibodies of the invention. The animal can be any farm animal or any animal raised for its meat, e.g., a pig, goat, cattle, sheep, horse and the like.

The animal feed additive of the invention may be a granulated enzyme product that may readily be mixed with feed components. Alternatively, feed additives of the invention can form a component of a pre-mix. The granulated enzyme product of the invention may be coated or uncoated. The particle size of the enzyme granulates can be compatible with that of feed and pre-mix components. This provides a safe and convenient mean of incorporating enzymes into feeds. Alternatively, the animal feed additive of the invention may be a stabilized liquid composition. This may be an aqueous or oil-based slurry. See, e.g., U.S. Patent No. 6,245,546.

A polypeptide, enzyme, protein, e.g. structural or binding protein, of the present invention, in the modification of animal feed or a food, can process the food or feed either *in vitro* (by modifying components of the feed or food) or *in vivo*. Polypeptides of the invention can be added to animal feed or food compositions.

In one aspect, an enzyme of the invention has any of the following enzyme activities, or is added in combination with another enzyme, e.g., beta-galactosidases,

catalases, laccases, cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases, transglutaminases, pectin methylesterases, cellobiohydrolases and/or transglutaminases. These enzyme digestion products are more digestible by the animal. Thus, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can contribute to the available energy of the feed or food. Also, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can improve the digestibility and uptake of carbohydrate and non-carbohydrate feed or food constituents such as protein, fat and minerals.

In another aspect, a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention can be supplied by expressing the enzymes directly in transgenic feed crops (as, e.g., transgenic plants, seeds and the like), such as grains, cereals, corn, soy bean, rape seed, lupin and the like. As discussed above, the invention provides transgenic plants, plant parts and plant cells comprising a nucleic acid sequence encoding a polypeptide of the invention. In one aspect, the nucleic acid is expressed such that the polypeptide, enzyme, protein, e.g. structural or binding protein of the invention is produced in recoverable quantities. The polypeptide, enzyme, protein, e.g. structural or binding protein can be recovered from any plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide can be used as such for improving the quality of a food or feed, e.g., improving nutritional value, palatability, etc.

The enzyme delivery matrix of the invention is in the form of discrete plural particles, pellets or granules. By "granules" is meant particles that are compressed or compacted, such as by a pelletizing, extrusion, or similar compacting to remove water from the matrix. Such compression or compacting of the particles also promotes intraparticle cohesion of the particles. For example, the granules can be prepared by pelletizing the grain-based substrate in a pellet mill. The pellets prepared thereby are ground or crumbled to a granule size suitable for use as an adjuvant in animal feed. Since the matrix is itself approved for use in animal feed, it can be used as a diluent for delivery of enzymes in animal feed.

The polypeptide, enzyme, protein, e.g. structural or binding protein contained in the invention enzyme delivery matrix and methods is in one aspect thermostable polypeptide, enzyme, protein, e.g. structural or binding protein, as described herein, so as to resist inactivation of the polypeptide, enzyme, protein, e.g. structural or binding protein during manufacture where elevated temperatures and/or steam may be employed to prepare the palletized enzyme delivery matrix. During digestion of feed containing the invention enzyme delivery matrix, aqueous digestive fluids will cause release of the active enzyme. Other types of thermostable enzymes and nutritional supplements that are thermostable can also be incorporated in the delivery matrix for release under any type of aqueous conditions.

A coating can be applied to the invention enzyme matrix particles for many different purposes, such as to add a flavor or nutrition supplement to animal feed, to delay release of animal feed supplements and enzymes in gastric conditions, and the like. Or, the coating may be applied to achieve a functional goal, for example, whenever it is desirable to slow release of the enzyme from the matrix particles or to control the conditions under which the enzyme will be released. The composition of the coating material can be such that it is selectively broken down by an agent to which it is susceptible (such as heat, acid or base, enzymes or other chemicals). Alternatively, two or more coatings susceptible to different such breakdown agents may be consecutively applied to the matrix particles.

The invention is also directed towards a process for preparing an enzyme-releasing matrix. In accordance with the invention, the process comprises providing discrete plural particles of a grain-based substrate in a particle size suitable for use as an enzyme-releasing matrix, wherein the particles comprise a polypeptide, enzyme, protein, e.g. structural or binding protein, encoded by an amino acid sequence of the invention. In one aspect, the process includes compacting or compressing the particles of enzyme-releasing matrix into granules, which most in one aspect is accomplished by pelletizing. The mold inhibitor and cohesiveness agent, when used, can be added at any suitable time, and in one aspect are mixed with the grain-based substrate in the desired proportions prior to pelletizing of the grain-based substrate. Moisture content in the pellet mill feed in one aspect is in the ranges set forth above with respect to the moisture content in the finished product, and in one aspect is about 14-15%. In one aspect, moisture is added to the feedstock in the form of an aqueous preparation of the enzyme to bring the feedstock to this moisture content. The temperature in the pellet mill in one aspect is brought to about

82°C with steam. The pellet mill may be operated under any conditions that impart sufficient work to the feedstock to provide pellets. The pelleting process itself is a cost-effective process for removing water from the enzyme-containing composition.

The compositions and methods of the invention can be practiced in conjunction with various nutritional and environmental factors including, e.g., (1) manipulation of gut microflora by supplementing feed with prebiotics and/or antibiotics, (2) low fiber diet (low energy and low purine diet), (3) restricting feed for 48 hours and withholding feed for 12 hours before slaughter, (4) increasing consumption of water, and/or (5) keeping animals clean.

The compositions and methods of the invention can be practiced in conjunction with administration of prebiotics, which are high molecular weight sugars, e.g., fructo-oligosaccharides (FOS); galacto-oligosaccharides (GOS), GRAS (Generally Recognized As Safe) material. These prebiotics can be metabolized by some probiotic lactic acid bacteria (LAB). They are non-digestible by the majority of intestinal microbes.

*Treating foods and food processing*

The polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention, e.g., as described in Tables 1, 2, and 3, have numerous applications in food processing industry. The invention provides treatment compositions, including, e.g., a plant cell, a bacterial cell, a yeast cell, an insect cell, or an animal cell, or any plant or plant part, or any food or feed, a waste product and the like.

The invention provides feeds or foods comprising a polypeptide, enzyme, protein, e.g. structural or binding protein, the invention, e.g., a feed, a liquid, e.g., a beverage (such as a fruit juice or a beer), a bread or a dough or a bread product, or a beverage precursor (e.g., a wort).

The food treatment processes of the invention can comprise use of any enzyme of the invention, which can have the following enzymatic activities, and also include the use of any combination of any enzyme, including lyases, laccases, catalases, laccases, cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases,

polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases, transglutaminases, pectin methylesterases, cellobiohydrolases and/or transglutaminases.

*Confectionaries, cacao butter and foods*

In one aspect, the compositions and methods of the invention can be used to make and process hard butters, such as cacao butter (cocao butter). The compositions and methods of the invention can be used to make cocoa butter alternatives by "structured" synthetic techniques using the enzymes of the invention, e.g., as described in Tables 1, 2, and 3, including esterases, acylases, lipases, phospholipases or proteases of the invention. For example, in one aspect, the methods of the invention process or synthesize triacylglycerides, diacylglycerides and/or monoacylglycerides for use as, e.g., cocoa butter alternatives. In one aspect, the methods of the invention generate a hard butter with a defined "plastic region" to maintain sufficient hardness below or at room temperature. In one aspect, the processed or synthesized lipid is designed to have a very narrow "plastic region," e.g., in one aspect, where it rapidly melts at about body temperature. Natural cacao butter begins to soften at approximately $30^0$C to $32^0$C, and completely melts at approximately $36^0$C. Natural cacao butter can contain 70 wt % or more of three 1,3-disaturated –2-oleoyl glycerols, which are 1,3-dipalmitoyl-2-oleoyl glycerol (POP), 1-palmitoyl-2-oleoyl glycerol (POSt) and 1,3-distearoyl-2-oleoyl glycerol (StOSt). These three glycerols show a similar melting behavior to each other and are responsible for melting properties of the cacao butter, exhibiting a very narrow plastic region. The invention provides synthetic cacao butters or processed cacao butters (synthesized or processed using a hydrolase of the invention, all possible composition are referred to as cocoa-butter alternatives) with varying percentages of ,3-dipalmitoyl-2-oleoyl glycerol (POP), 1-palmitoyl-2-oleoyl glycerol (POSt) and 1,3-distearoyl-2-oleoyl glycerol (StOSt), depending on the desired properties of the synthetic cacao butter, and, synthetic cacao butters with more or less than 70 wt % of the three 1,3-disaturated –2-oleoyl glycerols. The synthetic cacao butters of the invention can partially or completely replace natural or unprocessed cacao butters and can maintain or improve essential hard butter properties.

The invention provides synthetic cacao butters or processed cacao butters (synthesized or processed using a hydrolase of the invention) with desired properties for use in confectionary, bakery and pharmaceutical products. In one aspect, the invention provides confectionary, bakery and pharmaceutical products comprising a hydrolase of the invention. In one aspect, the methods of the invention make or process a lipid (a fat)

from a confection (e.g., a chocolate) or to be used in a confection. In one aspect, a lipid is made or processed such that the chocolate shows less finger-imprinting than chocolate made from natural cocoa butter, while still having sharp melting characteristics in the mouth. In one aspect, a lipid is made or processed such that a confection (e.g., chocolate) can be made at a comparatively high ambient temperature, or, be made using a cooling water at a comparatively high temperature. In one aspect, the lipid is made or processed such that a confection (e.g., chocolate) can be stored under relatively warmer conditions, e.g., tropical or semi-tropical conditions or in centrally heated buildings. In one aspect, the lipids are made or processed such that a confection (e.g., chocolate) will have a lipid (fat) content of consistent composition and quality. The enzymes of the invention can be used to provide a substitute composition for cacao butter which can significantly improve its thermal stability and replace it in a wide range of applications.

*Margarine and shortening production*

The invention provides synthetic or processed fats, e.g., margarine and shortening synthesized or processed using an enzyme of the invention, e.g., as described in Tables 1, 2, and 3, such as a hydrolase of the invention. In one aspect, the invention provides processed fats comprising a vegetable oil, such as soybean oil, corn oil, rapeseed oil, palm oil or lauric type oils synthesized or processed using a hydrolase of the invention. The synthetic or processed fats, e.g., margarine and shortening, are designed to have a desired "plasticity." Many of the plastic fat products, such as margarine and shortening, are produced from hard stocks and liquid oils as raw materials. For example, liquid oils such as soybean oil, corn oil, palm oil and rapeseed oil, are blended with their hardened oils (hard stocks), and the blend is adjusted to have an appropriate consistency (plasticity). The plastic fat products such as margarine and shortening so produced tend to cause the formation of relatively coarse crystallines because fats and oils used as the raw materials are composed of fatty acids having almost the same carbon chain length. In other words, they have a highly-unified composition of fatty acids. For this reason, the plasticity of these products can be maintained at an appropriate degree only within a narrow temperature range, so that the liquid oils contained therein have a tendency to exude. In one aspect, the invention provides methods of making or processing fats designed such that they have a varied (and defined) composition of fatty acids. The resultant oil, e.g., margarine or shortening, can have a broader range of plasticity.

In one aspect, the methods and compositions of the invention are used to make or process vegetable oils, such as soybean oil, corn oil, rapeseed oil, palm oil or lauric type

oils using the hydrolases of the invention, including inter-esterification and enzymatic transesterification, see e.g., U.S. Patent No. 5,288,619. The methods and compositions of the invention can be used in place of random inter-esterification as described in, e.g., U.S. Patent No. 3,949,105. In one aspect, the methods and compositions of the invention are used to in enzymatic transesterification for preparing an oil, e.g., a margarine oil, having both low trans- acid and low intermediate chain fatty acid content.

In one aspect, the symmetric structure of an oil, e.g., a palm or lauric type oils is modified, e.g., into a random structure. Thus, the methods of the invention can be used to modify the properties of plastic fat products. In one aspect, the modification of oils by the methods of the invention can be designed to prevent or slow gradually hardening of the oil with time, particularly when the products are being stored.

In one aspect, the methods and compositions of the invention in a trans-esterification reaction mixture comprising a stearic acid source material and an edible liquid vegetable oil, trans-esterifying the stearic acid source material and the vegetable oil using a 1-, 3-positionally specific lipase of the invention, and then hydrogenating the fatty acid mixture to provide a recycle stearic acid source material for a recyclic reaction with the vegetable oil. See e.g., U.S. Patent No. 5,288,619.

In one aspect, an inter-esterification reaction is conducted with a lipase of the invention. In one aspect, the lipase of the invention has a selectivity for the 1- and 3-positions of triglyceride to slow or inhibit an increase in the amount of tri-saturated triglycerides in the oil. In this reaction of the invention, deficiencies of conventional random inter-esterification and the difficulty of inter-esterification with a non-specific lipase can be overcome because the inter-esterification is conducted by an enzyme of the invention having a specificity for the 1- and 3- positions of triglycerides. In one aspect, the exudation of liquid oils contained in the products is slowed or prevented with a temperature increase in the reaction to inhibit a rise in the melting point caused by an increase in the amount of tri-saturated triglycerides. This addresses the problem of hardening of products during long-term storage.

*Brewing and fermenting*

The invention provides methods of brewing (e.g., fermenting) beer comprising hydrolases of the invention. In one exemplary process, starch-containing raw materials are disintegrated and processed to form a malt. A hydrolase of the invention is used at any point in the fermentation process. For example, hydrolases (e.g., proteases) of the invention can be used in the processing of barley malt. The major raw material of beer

brewing is barley malt. This can be a three stage process. First, the barley grain can be steeped to increase water content, e.g., to around about 40%. Second, the grain can be germinated by incubation at 15 to 25°C for 3 to 6 days when enzyme synthesis is stimulated under the control of gibberellins. In one aspect, hydrolases of the invention are added at this (or any other) stage of the process. The action of hydrolases results in an increase in fermentable reducing sugars. This can be expressed as the diastatic power, DP, which can rise from around 80 to 190 in 5 days at 12°C. Hydrolases (e.g., proteases) of the invention can be used in any beer or alcoholic beverage producing process, as described, e.g., in U.S. Patent No. 5,762,991; 5,536,650; 5,405,624; 5,021,246; 4,788,066.

*Waste treatment*

The polypeptide, enzyme, protein, e.g. structural or binding proteins of the invention, e.g., as described in Tables 1, 2, and 3, can be used in a variety of other industrial applications, e.g., in waste treatment (in addition to, e.g., biomass conversion to a fuel, e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel, and the like). For example, in one aspect, the invention provides a solid waste digestion process using a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention. The methods can comprise reducing the mass and volume of substantially untreated solid waste. Solid waste can be treated with an enzymatic digestive process in the presence of an enzymatic solution (including a polypeptide, enzyme, protein, e.g. structural or binding protein, of the invention) at a controlled temperature. This results in a reaction without appreciable bacterial fermentation from added microorganisms. The solid waste is converted into a liquefied waste and any residual solid waste. The resulting liquefied waste can be separated from said any residual solidified waste. See e.g., U.S. Patent No. 5,709,796.

In one aspect, the compositions and methods of the invention are used for odor removal or odor reduction in animal waste lagoons, e.g., on swine farms, and other animal waste management systems.

The waste treatment processes of the invention can comprise use of any enzyme of the invention, which can have the following enzymatic activities, and also include the use of any combination of any enzyme or protein, including e.g. structural or binding protein, catalases, lyases, laccases, cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases,

amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, phytases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases, transglutaminases, pectin methylesterases, cellobiohydrolases and/or transglutaminases.

*Increasing the flow of production fluids from a subterranean formation*

The invention also includes a method using an enzyme of the invention, e.g., as described in Tables 1, 2, and 3, wherein the method increases the flow of production fluids from a subterranean formation by removing viscous (e.g., starch-containing) damaging fluids formed during production operations; these fluids can be found within the subterranean formation which surrounds a completed well bore. Thus, this method of the invention results in production fluids being able to flow from the well bore. This method of the invention also addresses the problem of damaging fluids reducing the flow of production fluids from a formation below expected flow rates. In one aspect, the invention provides for formulating an enzyme treatment (using an enzyme of the invention) by blending together an aqueous fluid and a polypeptide of the invention; pumping the enzyme treatment to a desired location within the well bore; allowing the enzyme treatment to degrade the viscous (e.g., starch-containing) damaging fluid, whereby the fluid can be removed from the subterranean formation to the well surface; and wherein the enzyme treatment is effective to attack linkages in the viscous (e.g., starch-containing) fluid.

The subterranean formation enzyme treatment processes of the invention can comprise use of any enzyme of the invention, which can have the following enzymatic activities, and also include the use of any combination of any enzyme, including tryptophanases or tyrosine decarboxylases, laccases, catalases, lyases, laccases, other cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, other glucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases,

galactanases, pectin lyases, transglutaminases, pectin methylesterases, other cellobiohydrolases and/or transglutaminases.

*Latex processing*

The methods and compositions (e.g., enzymes of the invention, e.g., as described in Tables 1, 2, and 3, including esterases, acylases, lipases, phospholipases or proteases of the invention) of the invention can be used to selectively hydrolyze saturated esters over unsaturated esters into acids or alcohols. In one aspect, the invention provides for the selective hydrolysis of ethyl propionate over ethyl acrylate. In one aspect, these methods are used to remove undesired esters from monomer feeds used in latex polymerization and from the latexes after polymerization. The methods and compositions (hydrolases) of the invention can be used to treat latexes for a variety of purposes, e.g., to treat latexes used in hair fixative compositions to remove unpleasant odors. Latexes treated by the methods and compositions of the invention include, e.g., polymers containing acrylic, vinyl and unsaturated acid monomers, including alkyl acrylate monomers such as methyl acrylate, ethyl acrylate, propyl acrylate and butyl acrylate, and acrylate acids such as acrylic acid, methacrylic acid, crotonic acid, itaconic acid and mixtures thereof. See, e.g., U.S. Patent No. 5,856,150.

*Biomass conversion and production of clean bio fuels*

The invention provides enzymes, e.g., as described in Tables 1, 2, and 3 (including mixtures, or "cocktails" of enzymes) and methods for the conversion of a biomass or any lignocellulosic material (e.g., any composition comprising cellulose, hemicellulose and lignin), to a fuel (e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, biodiesel), in addition to feeds, foods and chemicals. Thus, the compositions and methods of the invention provide effective and sustainable alternatives or adjuncts to use of petroleum-based products, e.g., as a mixture of a biofuel such as biomethanol, bioethanol, biopropanol, biobutanol, and the like, to diesel fuel, gasoline, kerosene and the like. The invention provides organisms expressing enzymes of the invention for participation in chemical cycles involving natural biomass conversion. In one aspect, enzymes and methods for the conversion are used in enzyme ensembles for the efficient depolymerization of cellulosic and hemicellulosic polymers to metabolizeable carbon moieties. The invention provides methods for discovering and implementing the most effective of enzymes to enable these important new "biomass conversion" and alternative energy industrial processes.

The methods of the invention also include taking the converted lignocellulosic material (processed by enzymes of the invention) and making it into a fuel (e.g. a bioethanol, a biodiesel) by fermentation and/or by chemical synthesis. In one aspect, the produced sugars are fermented and/or the non-fermentable products are gasified.

The enzymes of the invention (including, for example, organisms, such as microorganisms, e.g., fungi, yeast or bacteria, making and in some aspects secreting recombinant enzymes of the invention) can be used in or included/ integrated at any stage of any biomass conversion process, e.g., at any one step, several steps, or included in all of the steps, or all of the following methods of biomass conversion processes, or all of these biofuel alternatives:

- Direct combustion: the burning of material by direct heat and is the simplest biomass technology; can be very economical if a biomass source is nearby.
- Pyrolysis: is the thermal degradation of biomass by heat in the absence of oxygen. In one aspect, biomass is heated to a temperature between about 800 and 1400 degrees Fahrenheit, but no oxygen is introduced to support combustion resulting in the creation of gas, fuel oil and charcoal.
- Gasification: biomass can be used to produce methane through heating or anaerobic digestion. Syngas, a mixture of carbon monoxide and hydrogen, can be derived from biomass.
    - Landfill Gas: is generated by the decay (anaerobic digestion) of buried garbage in landfills. When the organic waste decomposes, it generates gas consisting of approximately 50% methane, the major component of natural gas.
- Anaerobic digestion: converts organic matter to a mixture of methane, the major component of natural gas, and carbon dioxide. In one aspect, biomass such as waterwaste (sewage), manure, or food processing waste, is mixed with water and fed into a digester tank without air.
- Fermentation
    - Alcohol Fermentation: fuel alcohol is produced by converting starch to sugar, fermenting the sugar to alcohol, then separating the alcohol water mixture by distillation. Feedstocks such as wheat, barley, potatoes, and waste paper, sawdust, and straw containing sugar, starch, or cellulose can be converted to alcohol by fermentation with yeast.
- Transesterification: An exemplary reaction for converting oil to biodiesel is called transesterification. The transesterification process reacts an alcohol (like methanol) with the triglyceride oils contained in vegetable oils, animal fats, or recycled greases, forming fatty acid alkyl esters (biodiesel) and glycerin. The reaction requires heat and a strong base catalyst, such as sodium hydroxide or potassium hydroxide.
    - Biodiesel: Biodiesel is a mixture of fatty acid alkyl esters made from vegetable oils, animal fats or recycled greases. Biodiesel can be used as a fuel for vehicles in its pure form, but it is usually used as a petroleum diesel additive to reduce levels of particulates, carbon monoxide, hydrocarbons and air toxics from diesel-powered vehicles.

- Hydrolysis: includes hydrolysis of a compound, e.g., a biomass, such as a lignocellulosic material, catalyzed using an enzyme of the instant invention.
- Congeneration: is the simultaneous production of more than one form of energy using a single fuel and facility. In one aspect, biomass cogeneration has more potential growth than biomass generation alone because cogeneration produces both heat and electricity.

In one aspect, the polypeptides of the invention have cellulolytic activity, e.g., cellulases activity, such as endoglucanase, cellobiohydrolase and/or β-glucosidase (beta-glucosidase) activity, or other enzymatic activity for generating bioethanol, biopropanol, biobutanol, biopropanol, biomethanol and/or biodiesel from an organic material, e.g., a biomass, such as compositions derived from plants and animals, including any agricultural crop or other renewable feedstock, an agricultural residue or an animal waste, or the organic components of municipal and industrial wastes, or microorganisms such as algae or yeast. In one aspect, polypeptides of the invention are used in processes for converting lignocellulosic biomass to bioethanol, biopropanol, biobutanol, biopropanol, biomethanol and/or biodiesel, or otherwise are used in processes for hydrolyzing or digesting biomaterials such that they can be used as a biofuel (including bioethanol, biopropanol, biobutanol, biopropanol, biomethanol and/or biodiesel), or for making it easier for the biomass to be processed into a fuel. In an alternative aspect, polypeptides of the invention are used in processes for a transesterification process reacting an alcohol (like ethanol, propanol, butanol, propanol, methanol) with a triglyceride oil contained in a vegetable oil, animal fat or recycled greases, forming fatty acid alkyl esters (biodiesel) and glycerin. In one aspect, biodiesel is made from soybean oil or recycled cooking oils. Animal's fats, other vegetable oils, and other recycled oils can also be used to produce biodiesel, depending on their costs and availability. In another aspect, blends of all kinds of fats and oils are used to produce a biodiesel fuel of the invention.

Enzymes of the invention can also be used in glycerin refining. The glycerin by-product contains unreacted catalyst and soaps that are neutralized with an acid. Water and alcohol are removed to produce 50% to 80% crude glycerin. The remaining contaminants include unreacted fats and oils, which can be processes using the polypeptides of the invention. In a large biodiesel plants of the invention, the glycerin can be further purified, e.g., to 99% or higher purity, for the pharmaceutical and cosmetic industries.

Bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, and/or biodiesel made using the polypeptides of the invention can be used with fuel oxygenates to improve

combustion characteristics. Adding oxygen results in more complete combustion, which reduces carbon monoxide emissions. This is another environmental benefit of replacing petroleum fuels with biofuels (e.g., a fuel of the invention). A bioethanol made using the compositions and/or methods of this invention can be blended with gasoline to form an E10 blend (about 5% to 10% ethanol and about 90% to 95% gasoline), but it can be used in higher concentrations such as E85 or in its pure form. A bioethanol made using the compositions and/or methods of this invention can be blended with petroleum diesel to form a B20 blend (20% biodiesel and 80% petroleum diesel), although other blend levels can be used up to B100 (pure biodiesel).

The invention also provides processes for making ethanol ("bioethanol"), butanol ("biobutanol"), propanol ("biopropanol"), methanol ("biomethanol"), and/or diesel ("biodiesel") from compositions comprising any biomass, e.g., a lignocellulosic biomass. The biomass, e.g., lignocellulosic biomass, material can be obtained from agricultural crops, as a byproduct of food or feed production, or as lignocellulosic waste products, such as plant residues and waste paper, from algae, and the like. Examples of suitable plant or algae sources or plant or algae residues for treatment with polypeptides of the invention include kelp, various algae, grains, seeds, stems, leaves, hulls, husks, corn cobs, corn stover, straw, grasses (e.g., Indian grass, such as *Sorghastrum nutans*; or, switch grass, e.g., *Panicum* species, such as *Panicum virgatum*), and the like, as well as wood, wood chips, wood pulp, and sawdust. Examples of paper waste suitable for treatment with polypeptides of the invention include discard photocopy paper, computer printer paper, notebook paper, notepad paper, typewriter paper, and the like, as well as newspapers, magazines, cardboard, and paper-based packaging materials.

In one embodiment, the enzymes and methods of the invention can be used in conjunction with more "traditional" means of making ethanol, methanol, propanol, butanol, propanol and/or diesel from biomass, e.g., as methods comprising hydrolyzing lignocellulosic materials by subjecting dried lignocellulosic material in a reactor to a catalyst comprised of a dilute solution of a strong acid and a metal salt; this can lower the activation energy, or the temperature, of cellulose hydrolysis to obtain higher sugar yields; see, e.g., U.S. Patent Nos. 6,660,506 and 6,423,145.

Another exemplary method that incorporated use of enzymes of the invention comprises hydrolyzing lignocellulosic material containing hemicellulose, cellulose and lignin by subjecting the material to a first stage hydrolysis step in an aqueous medium at a temperature and a pressure chosen to effect primarily depolymerization of hemicellulose

without major depolymerization of cellulose to glucose. This step results in a slurry in which the liquid aqueous phase contains dissolved monosaccharides resulting from depolymerization of hemicellulose and a solid phase containing cellulose and lignin. A second stage hydrolysis step can comprise conditions such that at least a major portion of the cellulose is depolymerized, such step resulting in a liquid aqueous phase containing dissolved/ soluble depolymerization products of cellulose. See, e.g., U.S. Patent No. 5,536,325. Enzymes of the invention (including the invention's mixtures, or "cocktails" of enzymes) can be added at any stage of this exemplary process.

Another exemplary method that incorporated use of enzymes of the invention comprises processing a lignocellulose-containing biomass material by one or more stages of dilute acid hydrolysis with about 0.4% to 2% strong acid; and treating an unreacted solid lignocellulosic component of the acid hydrolyzed biomass material by alkaline delignification to produce precursors for biodegradable thermoplastics and derivatives. See, e.g., U.S. Patent No. 6,409,841. Enzymes of the invention can be added at any stage of this exemplary process.

Another exemplary method that incorporated use of enzymes of the invention comprises prehydrolyzing lignocellulosic material in a prehydrolysis reactor; adding an acidic liquid to the solid lignocellulosic material to make a mixture; heating the mixture to reaction temperature; maintaining reaction temperature for time sufficient to fractionate the lignocellulosic material into a solubilized portion containing at least about 20% of the lignin from the lignocellulosic material and a solid fraction containing cellulose; removing a solubilized portion from the solid fraction while at or near reaction temperature wherein the cellulose in the solid fraction is rendered more amenable to enzymatic digestion; and recovering a solubilized portion. See, e.g., U.S. Patent No. 5,705,369. Enzymes of the invention can be added at any stage of this exemplary process.

The invention provides methods for making motor fuel compositions (e.g., for spark ignition motors) based on liquid hydrocarbons blended with a fuel grade alcohol made by using an enzyme or a method of the invention. In one aspect, the fuels made by use of an enzyme of the invention comprise, e.g., coal gas liquid- or natural gas liquid-ethanol blends. In one aspect, a co-solvent is biomass-derived 2-methyltetrahydrofuran (MTHF). See, e.g., U.S. Patent No. 6,712,866.

In one aspect, methods of the invention for the enzymatic degradation of lignocellulose, e.g., for production of ethanol from lignocellulosic material, can also

comprise use of ultrasonic treatment of the biomass material; see, e.g., U.S. Patent No. 6,333,181.

In another aspect, methods of the invention for producing bioethanol, biopropanol, biobutanol, biopropanol, biomethanol, and/or biodiesel from a cellulosic substrate comprise providing a reaction mixture in the form of a slurry comprising cellulosic substrate, an enzyme of this invention and a fermentation agent (e.g., within a reaction vessel, such as a semi-continuously solids-fed bioreactor), and the reaction mixture is reacted under conditions sufficient to initiate and maintain a fermentation reaction (as described, e.g., in U.S. Pat. App. No. 20060014260). In one aspect, experiment or theoretical calculations can determine an optimum feeding frequency. In one aspect, additional quantities of the cellulosic substrate and the enzyme are provided into the reaction vessel at an interval(s) according to the optimized feeding frequency.

One exemplary process for making a biofuel (e.g., bioethanol, biopropanol, biobutanol, biopropanol, biomethanol) and/or biodiesel of the invention is described in U.S. Pat. App. Pub. Nos. 20050069998; 20020164730; and in one aspect comprises stages of grinding the lignocellulosic biomass (e.g., to a size of 15-30 mm), subjecting the product obtained to steam explosion pre-treatment (e.g., at a temperature of 190-230°C) for between 1 and 10 minutes in a reactor; collecting the pre-treated material in a cyclone or related product of manufacture; and separating the liquid and solid fractions by filtration in a filter press, introducing the solid fraction in a fermentation deposit and adding one or more enzymes of the invention, e.g., a cellulase and/or beta-glucosidase enzyme (e.g., dissolved in citrate buffer pH 4.8).

Another exemplary process for making a biofuels (such as bioethanol, biopropanol, biobutanol, biopropanol, biomethanol) and/or biodiesels of the invention comprising ethanol using enzymes of the invention comprises pretreating a starting material comprising a lignocellulosic feedstock comprising at least hemicellulose and cellulose. In one aspect, the starting material comprises potatoes, soybean (rapeseed), barley, rye, corn, oats, wheat, beets or sugar cane or a component or waste or food or feed production byproduct. The starting material ("feedstock") is reacted at conditions which disrupt the plant's fiber structure to effect at least a partial hydrolysis of the hemicellulose and cellulose. Disruptive conditions can comprise, e.g., subjecting the starting material to an average temperature of 180°C to 270°C at pH 0.5 to 2.5 for a period of about 5 seconds to 60 minutes; or, temperature of 220°C to 270°C, at pH 0.5 to 2.5 for a period of 5 seconds to 120 seconds, or equivalent. This generates a feedstock with increased

accessibility to being digested by an enzyme, e.g., a cellulase enzyme of the invention. U.S. Patent No. 6,090,595.

Exemplary conditions for cellulase hydrolysis of lignocellulosic material include reactions at temperatures between about 30°C and 48°C, and/or a pH between about 4.0 and 6.0. Other exemplary conditions include a temperature between about 30°C and 60°C and a pH between about 4.0 and 8.0.

*Detergent Compositions*

The invention provides detergent compositions comprising one or more polypeptides of the invention (e.g., enzymes as described in Tables 1, 2, and 3, e.g., having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity) and methods of making and using these compositions. The invention incorporates all methods of making and using detergent compositions, see, e.g., U.S. Patent No. 6,413,928; 6,399,561; 6,365,561; 6,380,147. The detergent compositions can be a one and two part aqueous composition, a non-aqueous liquid composition, a cast solid, a granular form, a particulate form, a compressed tablet, a gel and/or a paste and a slurry form. The invention also provides methods capable of a rapid removal of gross food soils, films of food residue and other minor food compositions using these detergent compositions. Enzymes of the invention can facilitate the removal of starchy stains by means of catalytic hydrolysis of the starch polysaccharide. Enzymes of the invention can be used in dishwashing detergents in textile laundering detergents.

The actual active enzyme content depends upon the method of manufacture of a detergent composition and is not critical, assuming the detergent solution has the desired enzymatic activity. In one aspect, the amount of glucosidase present in the final solution ranges from about 0.001 mg to 0.5 mg per gram of the detergent composition. The particular enzyme chosen for use in the process and products of this invention depends upon the conditions of final utility, including the physical product form, use pH, use temperature, and soil types to be degraded or altered. The enzyme can be chosen to provide optimum activity and stability for any given set of utility conditions. In one aspect, the polypeptides of the present invention are active in the pH ranges of from about 4 to about 12 and in the temperature range of from about 20°C to about 95°C. The detergents of the invention can comprise cationic, semi-polar nonionic or zwitterionic surfactants; or, mixtures thereof.

Enzymes of the present invention (e.g., enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity) can be formulated into

powdered and liquid detergents having pH between 4.0 and 12.0 at levels of about 0.01 to about 5% (preferably 0.1% to 0.5%) by weight. These detergent compositions can also include other enzymes such as known proteases, cellulases, lipases or endoglycosidases, as well as builders and stabilizers. The addition of enzymes of the invention to conventional cleaning compositions does not create any special use limitation. In other words, any temperature and pH suitable for the detergent is also suitable for the present compositions as long as the pH is within the above range, and the temperature is below the described enzyme's denaturing temperature. In addition, the polypeptides of the invention can be used in a cleaning composition without detergents, again either alone or in combination with builders and stabilizers.

The present invention provides cleaning compositions including detergent compositions for cleaning hard surfaces, detergent compositions for cleaning fabrics, dishwashing compositions, oral cleaning compositions, denture cleaning compositions, and contact lens cleaning solutions.

In one aspect, the invention provides a method for washing an object comprising contacting the object with a polypeptide of the invention under conditions sufficient for washing. A polypeptide of the invention may be included as a detergent additive. The detergent composition of the invention may, for example, be formulated as a hand or machine laundry detergent composition comprising a polypeptide of the invention. A laundry additive suitable for pre-treatment of stained fabrics can comprise a polypeptide of the invention. A fabric softener composition can comprise a polypeptide of the invention. Alternatively, a polypeptide of the invention can be formulated as a detergent composition for use in general household hard surface cleaning operations. In alternative aspects, detergent additives and detergent compositions of the invention may comprise one or more other enzymes such as a protease, a lipase, a cutinase, another glucosidase, a carbohydrase, another cellulase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, e.g., a lactase, and/or a peroxidase. The properties of the enzyme(s) of the invention are chosen to be compatible with the selected detergent (i.e. pH-optimum, compatibility with other enzymatic and non-enzymatic ingredients, etc.) and the enzyme(s) is present in effective amounts. In one aspect, enzymes of the invention are used to remove malodorous materials from fabrics. Various detergent compositions and methods for making them that can be used in practicing the invention are described in, e.g., U.S. Patent Nos. 6,333,301; 6,329,333; 6,326,341; 6,297,038; 6,309,871; 6,204,232; 6,197,070; 5,856,164.

The detergents and related processes of the invention can also include the use of any combination of other enzymes such as tryptophanases or tyrosine decarboxylases, laccases, catalases, laccases, other cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, other glucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases, transglutaminases, pectin methylesterases, other cellobiohydrolases and/or transglutaminases.

*Treating fabrics and textiles*

The invention provides compositions and methods of treating fabrics and textiles using one or more polypeptides of the invention, e.g., enzymes as described in Tables 1, 2, and 3, including enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity. The polypeptides of the invention can be used in any fabric-treating method, which are well known in the art, see, e.g., U.S. Patent No. 6,077,316. For example, in one aspect, the feel and appearance of a fabric is improved by a method comprising contacting the fabric with an enzyme of the invention in a solution. In one aspect, the fabric is treated with the solution under pressure.

In one aspect, the enzymes of the invention are applied during or after the weaving of textiles, or during the desizing stage, or one or more additional fabric processing steps. During the weaving of textiles, the threads are exposed to considerable mechanical strain. Prior to weaving on mechanical looms, warp yarns are often coated with sizing starch or starch derivatives in order to increase their tensile strength and to prevent breaking. The enzymes of the invention can be applied to remove these sizing starch or starch derivatives. After the textiles have been woven, a fabric can proceed to a desizing stage. This can be followed by one or more additional fabric processing steps. Desizing is the act of removing size from textiles. After weaving, the size coating must be removed before further processing the fabric in order to ensure a homogeneous and wash-proof result. The invention provides a method of desizing comprising enzymatic hydrolysis of the size by the action of an enzyme of the invention.

The enzymes of the invention (e.g., enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity) can be used to desize

fabrics, including cotton-containing fabrics, as detergent additives, e.g., in aqueous compositions. The invention provides methods for producing a stonewashed look on indigo-dyed denim fabric and garments. For the manufacture of clothes, the fabric can be cut and sewn into clothes or garments, which is afterwards finished. In particular, for the manufacture of denim jeans, different enzymatic finishing methods have been developed. The finishing of denim garment normally is initiated with an enzymatic desizing step, during which garments are subjected to the action of amylolytic enzymes in order to provide softness to the fabric and make the cotton more accessible to the subsequent enzymatic finishing steps. The invention provides methods of finishing denim garments (e.g., a "bio-stoning process"), enzymatic desizing and providing softness to fabrics using the Enzymes of the invention. The invention provides methods for quickly softening denim garments in a desizing and/or finishing process.

The invention also provides disinfectants comprising enzymes of the invention (e.g., enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity).

The fabric or textile treatment processes of the invention can also include the use of any combination of other enzymes such as tryptophanases or tyrosine decarboxylases, laccases, catalases, laccases, other cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, other glucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases, transglutaminases, pectin methylesterases, other cellobiohydrolases and/or transglutaminases.

*Paper or pulp treatment*

The enzymes of the invention e.g., enzymes as described in Tables 1, 2, and 3, including enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity) can be in paper or pulp treatment or paper deinking. For example, in one aspect, the invention provides a paper treatment process using enzymes of the invention. In one aspect, the enzymes of the invention can be used to modify starch in the paper thereby converting it into a liquefied form. In another aspect, paper components of recycled photocopied paper during chemical and enzymatic deinking

processes. In one aspect, Enzymes of the invention can be used in combination with other enzymes, including other cellulases (including other endoglucanases, cellobiohydrolases and/or beta-glucosidases). The wood, paper, paper product or pulp can be treated by the following three processes: 1) disintegration in the presence of an enzyme of the invention, 2) disintegration with a deinking chemical and an enzyme of the invention, and/or 3) disintegration after soaking with an enzyme of the invention. The recycled paper treated with an enzyme of the invention can have a higher brightness due to removal of toner particles as compared to the paper treated with just cellulase. While the invention is not limited by any particular mechanism, the effect of an enzyme of the invention may be due to its behavior as surface-active agents in pulp suspension.

The invention provides methods of treating paper and paper pulp using one or more polypeptides of the invention. The polypeptides of the invention can be used in any paper- or pulp-treating method, which are well known in the art, see, e.g., U.S. Patent No. 6,241,849; 6,066,233; 5,582,681. For example, in one aspect, the invention provides a method for deinking and decolorizing a printed paper containing a dye, comprising pulping a printed paper to obtain a pulp slurry, and dislodging an ink from the pulp slurry in the presence of an enzyme of the invention (other enzymes can also be added). In another aspect, the invention provides a method for enhancing the freeness of pulp, e.g., pulp made from secondary fiber, by adding an enzymatic mixture comprising an enzyme of the invention (can also include other enzymes, e.g., pectinase enzymes) to the pulp and treating under conditions to cause a reaction to produce an enzymatically treated pulp. The freeness of the enzymatically treated pulp is increased from the initial freeness of the secondary fiber pulp without a loss in brightness.

The paper, wood or pulp treatment or recycling processes of the invention can also include the use of any combination of other enzymes such as tryptophanases or tyrosine decarboxylases, laccases, catalases, laccases, other cellulases, endoglycosidases, endo-beta-1,4-laccases, amyloglucosidases, other glucosidases, glucose isomerases, glycosyltransferases, lipases, phospholipases, lipooxygenases, beta-laccases, endo-beta-1,3(4)-laccases, cutinases, peroxidases, amylases, glucoamylases, pectinases, reductases, oxidases, decarboxylases, phenoloxidases, ligninases, pullulanases, arabinanases, hemicellulases, mannanases, xylolaccases, xylanases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, proteases, peptidases, proteinases, polygalacturonases, rhamnogalacturonases, galactanases, pectin lyases,

transglutaminases, pectin methylesterases, other cellobiohydrolases and/or transglutaminases.

*Repulping: treatment of lignocellulosic materials*

The invention also provides a method for the treatment of lignocellulosic fibers, wherein the fibers are treated with a polypeptide of the invention e.g., enzymes as described in Tables 1, 2, and 3, including enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity), in an amount which is efficient for improving the fiber properties. The enzymes of the invention may also be used in the production or recycling of lignocellulosic materials such as pulp, paper and cardboard, from starch reinforced waste paper and cardboard, especially where repulping or recycling occurs at pH above 7 and where the enzymes of the invention can facilitate the disintegration of the waste material through degradation of the reinforcing starch. The enzymes of the invention can be useful in a process for producing a papermaking pulp from starch-coated printed paper. The process may be performed as described in, e.g., WO 95/14807. An exemplary process comprises disintegrating the paper to produce a pulp, treating with a starch-degrading enzyme before, during or after the disintegrating, and separating ink particles from the pulp after disintegrating and enzyme treatment. See also U.S. Patent No. 6,309,871 and other US patents cited herein. Thus, the invention includes a method for enzymatic deinking of recycled paper pulp, wherein the polypeptide is applied in an amount which is efficient for effective de-inking of the fiber surface.

*Brewing and fermenting*

The invention provides compositions for and methods of brewing (e.g., fermenting) beer comprising an enzyme of the invention, e.g., enzymes as described in Tables 1, 2, and 3, including enzymes having cellulase, endoglucanase, cellobiohydrolase, mannanase and/or beta-glucosidase activity. In one exemplary process, starch-containing raw materials are disintegrated and processed to form a malt. An enzyme of the invention is used at any point in the fermentation process. For example, enzymes of the invention can be used in the processing of barley malt. The major raw material of beer brewing is barley malt. This can be a three stage process. First, the barley grain can be steeped to increase water content, e.g., to around about 40%. Second, the grain can be germinated by incubation at 15-25°C for 3 to 6 days when enzyme synthesis is stimulated under the control of gibberellins. During this time enzyme levels rise significantly. In one aspect,

enzymes of the invention are added at this (or any other) stage of the process. The action of the enzyme results in an increase in fermentable reducing sugars. This can be expressed as the diastatic power, DP, which can rise from around 80 to 190 in 5 days at 12°C.

Enzymes of the invention can be used in any beer producing process, as described, e.g., in U.S. Patent No. 5,762,991; 5,536,650; 5,405,624; 5,021,246; 4,788,066.

*Pharmaceutical compositions and dietary supplements*

The invention also provides pharmaceutical compositions and dietary supplements (e.g., dietary aids) comprising an enzyme of the invention (e.g., an exemplary enzyme of the invention, including those described in Tables 1, 2, and 3). In one aspect, the pharmaceutical compositions and dietary supplements (e.g., dietary aids) are formulated for oral ingestion, e.g., to improve the digestibility of foods and feeds having a high cellulose or lignocellulosic component. See also discussion, above.

Periodontal treatment compounds can comprise an enzyme of the invention, e.g., as described in U.S. patent no. 6,776,979. Compositions and methods for the treatment or prophylaxis of acidic gut syndrome can comprise an enzyme of the invention, e.g., as described in U.S. patent no. 6,468,964.

In another aspect, wound dressings, implants and the like comprise antimicrobial (e.g., antibiotic-acting) enzymes, including an enzyme of the invention (including, e.g., exemplary sequences of the invention). Enzymes of the invention can also be used in alginate dressings, antimicrobial barrier dressings, burn dressings, compression bandages, diagnostic tools, gel dressings, hydro-selective dressings, hydrocellular (foam) dressings, hydrocolloid dressings, I.V dressings, incise drapes, low adherent dressings, odor absorbing dressings, paste bandages, post operative dressings, scar management, skin care, transparent film dressings and/or wound closure. Enzymes of the invention can be used in wound cleansing, wound bed preparation, to treat pressure ulcers, leg ulcers, burns, diabetic foot ulcers, scars, IV fixation, surgical wounds and minor wounds. Enzymes of the invention can be used to in sterile enzymatic debriding compositions, e.g., ointments. In various aspects, the cellulase is formulated as a tablet, gel, pill, implant, liquid, spray, powder, food, feed pellet or as an encapsulated formulation.

The polypeptides of the invention, such as those having hydrolases activity (e.g., lipases, esterase, protease and/or phospholipases) of the invention can be used in detoxification processes, e.g., for the detoxification of endotoxins, e.g., compositions comprising lipopolysaccharides (LPS), and, the invention provides detoxification

processes using at least one enzyme of the invention. In one aspect, a lipase and/or an esterase of the invention is used to detoxify a lipopolysaccharide (LPS). In one aspect, this detoxification is by deacylation of 2' and/or 3' fatty acid chains from lipid A. In one aspect, a hydrolase (e.g., a lipase and/or an esterase) of the invention is used to hydrolyze a 2'-lauroyl and/or a 3'-myristoyl chain from a lipid, e.g., a lipid A (e.g., from a bacterial endotoxin). In one aspect, the process of the invention is used to destroy an endotoxin, e.g., a toxin from a gram negative bacteria, as from *E. coli*. In one aspect, a hydrolase (e.g., a lipase and/or an esterase) of the invention is used to ameliorate the effects of toxin poisoning (e.g., from an on-going gram negative infection), or, to prophylactically to prevent the effects of endotoxin during an infection (e.g., an infection in an animal or a human). Accordingly, the invention provides a pharmaceutical composition comprising a hydrolase (e.g., a lipase and/or an esterase) of the invention, and method using a hydrolase of the invention, for the amelioration or prevention of lipopolysaccharide (LPS) toxic effects, e.g., during sepsis.

*Biodefense applications*

In other aspects, an enzyme of the invention (e.g., an exemplary enzyme of the invention, including those described in Tables 1, 2, and 3) can be used in biodefense (e.g., destruction of spores or bacteria comprising a lignocellulosic material). Use of enzymes of the invention in biodefense applications offer a significant benefit, in that they can be very rapidly developed against any currently unknown or biological warfare agents of the future. In addition, enzymes of the invention can be used for decontamination of affected environments. In aspect, the invention provides a biodefense or bio-detoxifying agent comprising a polypeptide of the invention (including, e.g., exemplary sequences of the invention), or a polypeptide encoded by a nucleic acid of the invention (including, e.g., exemplary sequences of the invention).

*Nutraceuticals*

In one aspect, the compositions, e.g., an exemplary enzyme of the invention, including those described in Tables 1, 2, and 3, and methods of the invention can be used to make nutraceuticals by processing or synthesizing lipids and oils using the enzymes of the invention, e.g., esterases, acylases, lipases, phospholipases or proteases of the invention. In one aspect, the processed or synthesized lipids or oils include poly-unsaturated fatty acids (PUFAs), diacylglycerides, e.g., 1,3-diacyl glycerides (DAGs), monoacylglycerides, e.g., 2-monoacylglycerides (MAGs) and triacylglycerides (TAGs).

In one aspect, the nutraceuticals is made by processing diacylglycerides, e.g., 1,3-diacyl glycerides (DAGs), monoacylglycerides, e.g., 2-monoacylglycerides (MAGs) and/or triacylglycerides (TAGs) from plant (e.g., oilseed) sources or from animal (e.g., fish oil) sources.

In one aspect, the compositions and methods of the invention can be used to fortify dietary compositions, especially cow's milk based products, e.g., cow's milk-based infant formulas, with bile salt-activated hydrolases. The compositions made by the methods and compositions of the invention can be used to feed newborn and premature infants, including administration of a bile salt-activated hydrolase of the invention to increase fat digestion and therefore growth rate. Similarly, the invention provides compositions and methods for treating subjects for inadequate pancreatic enzyme production by administration of bile salt-activated hydrolase in conjunction with ingestion of fats; see also discussion, below.

In one aspect, the invention provides a dietary composition comprising a hydrolase of the invention, e.g., bile salt-activated hydrolase of the invention. In one aspect, the invention provides a dietary composition comprising a nutritional base comprising a fat and an effective amount of bile salt-activated hydrolase of the invention. In one aspect, the invention provides a cow's milk-based infant formula comprising a hydrolase of the invention, e.g., bile salt-activated hydrolase of the invention. In one aspect, the hydrolase of the invention is active in the digestion of long chain fatty acids, e.g., $C_{12}$ to $C_{22}$, which make up a very high percentage of most milks, e.g., 99% of human breast milk. See, e.g., U.S. Patent No. 5,000,975.

In one aspect, the invention provides a dietary composition comprising a vegetable oil fat and a hydrolase of the invention. The invention provides methods of processing milk based products and/or vegetable oil-comprising compositions to make dietary compositions. In one aspect, the processed compositions comprise a lauric acid oil, an oleic acid oil, a palmitic acid oil and/or a linoleic acid oil. In one aspect, a rice bran oil, sunflower oleic oil and/or canola oil may be used as oleic acids oils. In one aspect, fats and oils, e.g., oilseeds, from plants, including, e.g., rice, canola, sunflower, olive, palm, soy or lauric type oils for use in the nutraceuticals and dietary compositions are processed or made using a hydrolase of the invention. See, e.g., U.S. Patent No. 4,944,944.

In one aspect, the enzymes of the invention are provided in a form that is stable to storage in the formula and/or the stomach, but active when the formulation reaches the

portion of the gastrointestinal tract where the formula would normally be digested. Formulations (e.g., microcapsules) for release in the intestine are well known in the art, e.g., biodegradable polymers such as polylactide and polyglycolide, as described, e.g., in U.S. Patent. Nos. 4,767,628; 4,897,268; 4,925,673; 5,902,617.

The invention will be further described with reference to the following examples; however, it is to be understood that the invention is not limited to such examples.

## EXAMPLES

EXAMPLE 1: Proteomic Analysis of Exemplary Polypeptides of the Invention

The following example demonstrates a protocol used for proteomic analysis of exemplary polypeptides of the invention. All of the exemplary nucleic acid sequences were initially isolated from termite gut samples (termite hind gut fluid samples).

This proteomic analysis of exemplary polypeptides of the invention found small peptides of whole open reading frames (ORFs); where the ORFs were identified by even SEQ ID NO: in the first column of the Table 4, below (listing exemplary amino acid sequences of the invention). The greater the number of peptides that were identified (see third column of the table), the better the confidence that a correct ORF was identified (by sequence). The 2nd column (spectral count) is abundance - the number of times the peptides were found.

SAMPLE DIGESTION: The termite hind gut fluid samples were centrifuged for 10 min (4°C) at 13000 rpm using bench top microfuge to remove insoluble fraction of the fluid. 500 micro gram proteins from the soluble fraction was heat treated at 100°C for 5 min in the presence of 0.5% RAPIGEST™ (Waters Corporation, Milford, MA). The samples where reduced with 1 mM DTT and alkylated with 0.4 mg/mL of iodoacetamide. After digestion with endoproteinase Lys-C (1/200 of protein), the protein samples were further digested twice with trypsin (1/100 of protein). The second batch of trypsin was added 4 h after the addition of the first batch. The RAPIGEST™ was removed from the solution by addition of HCl to final concentration of 0.05M. Samples where incubated at 37°C for 60 minutes and centrifuged at 13000 rpm using bench top microfuge to remove insoluble fraction. The pH of the soluble fraction was adjusted using ammonium hydroxide to pH 2-3.

Online 3D LC-MS/MS. The digested termite protein samples were analyzed by the 3D LC-MS/MS system as described by Jing Wei et al (2005) J. Proteome Res. May-Jun;4(3):801-8. The LTQ mass spectrometer was set to divide the full MS scan into three smaller sections covering a total range of 400-2000 m/z. Each of the smaller MS scans was followed by 4 MS/MS scans of the most intense ions from the preceding MS scan. The typical collision energy for collision-induced dissociation was set to 35% with a 30-ms activation time. Dynamic exclusion was enabled with a repeat count of 1 and a 3-min exclusion duration window.

Data Analysis. The LC-MS/MS raw data were extracted using the XCALIBUR RAWFILE CONVERTER V 1.0.0a™ (Advanced Chemistry Development, Inc., (ACD/Labs), Toronto, Ontario, Canada) and then searched against the termite meta genome predicted open reading frames using the SEQUEST™ program (The Scripps Research Institute, La Jolla, CA). The nonspecific cleavage rule was designated during the SEQUEST™ search. Differential modifications of Met oxidation (+16) and Cys alkylation (+57) were allowed for the database search. The results were filtered using the same criteria set as previously described6,8 to obtain the peptide identifications. Briefly, all peptide identifications had a ÄCn > 0.1; peptides with a +1 charge state were fully tryptic with XCorr > 1.9; peptides with a +2 charge state were with XCorr > 3.0 or fully or partially tryptic with XCorr ranging between 2.2 and 3.0; peptides with a +3 charge state were fully or partially tryptic with XCorr > 3.5. The PORTFOLIO™ program was used for protein identification and summarizing the results. For each protein sequence in SGD, the program assembled the peptide identifications that matched its substring. The protein identification was established if one or more peptide identifications were matched. The peptide identifications were allowed for multiple protein identifications. The number of distinct peptides and sequence coverage per protein identification were also calculated. Proteins identified by a single peptide are considered at a reduced level of confidence. These protein identifications by single peptide were included in all results and their percentages in the total protein identifications were compared among all systems. The cumulative total of protein identifications was calculated as the sum of yeast sequences from SGD that were detected in at least one of the protein samples.

To aid in reading Table 5, below, for example, in the first row, the data confirms that the exemplary nucleic acid of the invention SEQ ID NO:10011 (encoding SEQ ID NO:10012) is expressed, and two unique peptides were confirmed by this analysis, and as

noted above, spectral count denotes abundance, or the number of times the peptides were found.

Table 5

| SEQ ID NO: | total spectral count | number of unique peptides |
|---|---|---|
| 10011, 10012 | 6 | 2 |
| 10017, 10018 | 1 | 1 |
| 10155, 10156 | 6 | 3 |
| 10233, 10234 | 2 | 1 |
| 10337, 10338 | 1 | 1 |
| 10419, 10420 | 1 | 1 |
| 10453, 10454 | 1 | 1 |
| 10467, 10468 | 4 | 2 |
| 10507, 10508 | 1 | 1 |
| 10569, 10570 | 1 | 1 |
| 1059, 1060 | 2 | 1 |
| 10595, 10596 | 1 | 1 |
| 10599, 10600 | 4 | 1 |
| 10653, 10654 | 1 | 1 |
| 1067, 1068 | 1 | 1 |
| 10679, 10680 | 90 | 6 |
| 10685, 10686 | 1 | 1 |
| 10697, 10698 | 1 | 1 |
| 1071, 1072 | 1 | 1 |
| 1073, 1074 | 24 | 5 |
| 10799, 10800 | 1 | 1 |
| 10815, 10816 | 1 | 1 |
| 10851, 10852 | 99 | 7 |
| 10867, 10868 | 4 | 2 |
| 10891, 10892 | 136 | 10 |
| 10895, 10896 | 1 | 1 |
| 10959, 10960 | 1 | 1 |
| 10965, 10966 | 1 | 1 |
| 10985, 10986 | 1 | 1 |
| 10993, 10994 | 1 | 1 |
| 10995, 10996 | 18 | 4 |
| 11055, 11056 | 4 | 1 |
| 11091, 11092 | 1 | 1 |
| 11199, 11200 | 15 | 2 |
| 11239, 11240 | 4 | 3 |
| 11307, 11308 | 89 | 7 |
| 11309, 11310 | 1 | 1 |
| 11335, 11336 | 3 | 1 |
| 11355, 11356 | 5 | 2 |
| 11377, 11378 | 2 | 1 |
| 11381, 11382 | 1 | 1 |
| 11425, 11426 | 1 | 1 |
| 11443, 11444 | 2 | 1 |
| 11449, 11450 | 1 | 1 |

| | | |
|---|---|---|
| 11587, 11588 | 2 | 1 |
| 11595, 11596 | 17 | 3 |
| 11623, 11624 | 1 | 1 |
| 11661, 11662 | 4 | 1 |
| 11681, 11682 | 1 | 1 |
| 11687, 11688 | 1 | 1 |
| 11699, 11700 | 1 | 1 |
| 11703, 11704 | 1 | 1 |
| 1171, 1172 | 16 | 5 |
| 11779, 11780 | 1 | 1 |
| 11783, 11784 | 1 | 1 |
| 11795, 11796 | 1 | 1 |
| 1183, 1184 | 1 | 1 |
| 11861, 11862 | 1 | 1 |
| 11865, 11866 | 1 | 1 |
| 11883, 11884 | 1 | 1 |
| 11929, 11930 | 1 | 1 |
| 11941, 11942 | 16 | 1 |
| 1197, 1198 | 10 | 4 |
| 11995, 11996 | 17 | 5 |
| 11999, 12000 | 2 | 1 |
| 12051, 12052 | 1 | 1 |
| 12069, 12070 | 1 | 1 |
| 12091, 12092 | 15 | 2 |
| 12099, 12100 | 1 | 1 |
| 12123, 12124 | 1 | 1 |
| 12155, 12156 | 10 | 1 |
| 12173, 12174 | 3 | 2 |
| 12203, 12204 | 1 | 1 |
| 1225, 1226 | 1 | 1 |
| 1229, 1230 | 43 | 4 |
| 1233, 1234 | 2 | 1 |
| 12347, 12348 | 8 | 4 |
| 12373, 12374 | 1 | 1 |
| 12387, 12388 | 7 | 1 |
| 12405, 12406 | 4 | 2 |
| 12409, 12410 | 2 | 2 |
| 12419, 12420 | 3 | 3 |
| 12447, 12448 | 1 | 1 |
| 12449, 12450 | 1 | 1 |
| 12455, 12456 | 1 | 1 |
| 12471, 12472 | 2 | 1 |
| 12493, 12494 | 80 | 8 |
| 1253, 1254 | 3 | 2 |
| 12543, 12544 | 1 | 1 |
| 12591, 12592 | 6 | 3 |
| 12617, 12618 | 1 | 1 |
| 12621, 12622 | 41 | 10 |
| 12703, 12704 | 1 | 1 |
| 12749, 12750 | 57 | 4 |
| 12765, 12766 | 1 | 1 |

| | | |
|---|---|---|
| 12883, 12884 | 18 | 4 |
| 12965, 12966 | 1 | 1 |
| 12971, 12972 | 7 | 1 |
| 1303, 1304 | 28 | 3 |
| 13075, 13076 | 22 | 4 |
| 13195, 13196 | 2 | 1 |
| 13241, 13242 | 1 | 1 |
| 13289, 13290 | 1 | 1 |
| 13299, 13300 | 1 | 1 |
| 13317, 13318 | 1 | 1 |
| 1337, 1338 | 1 | 1 |
| 13409, 13410 | 9 | 3 |
| 13441, 13442 | 2 | 1 |
| 13451, 13452 | 1 | 1 |
| 13481, 13482 | 1 | 1 |
| 1349, 1350 | 2 | 1 |
| 13529, 13530 | 10 | 3 |
| 13597, 13598 | 1 | 1 |
| 13605, 13606 | 2 | 2 |
| 1363, 1364 | 17 | 4 |
| 1365, 1366 | 55 | 2 |
| 13667, 13668 | 3 | 1 |
| 13675, 13676 | 1 | 1 |
| 13711, 13712 | 8 | 2 |
| 13747, 13748 | 1 | 1 |
| 13757, 13758 | 1 | 1 |
| 13795, 13796 | 146 | 7 |
| 13803, 13804 | 3 | 2 |
| 1381, 1382 | 1 | 1 |
| 13813, 13814 | 4 | 1 |
| 13835, 13836 | 2 | 1 |
| 13847, 13848 | 1 | 1 |
| 1387, 1388 | 2 | 1 |
| 13875, 13876 | 1 | 1 |
| 13915, 13916 | 1 | 1 |
| 14003, 14004 | 1 | 1 |
| 14011, 14012 | 1 | 1 |
| 14015, 14016 | 68 | 5 |
| 14057, 14058 | 2 | 1 |
| 14073, 14074 | 1 | 1 |
| 14079, 14080 | 2 | 2 |
| 14237, 14238 | 5 | 1 |
| 14261, 14262 | 4 | 1 |
| 14357, 14358 | 17 | 2 |
| 14411, 14412 | 4 | 1 |
| 14415, 14416 | 1 | 1 |
| 14423, 14424 | 3 | 1 |
| 14509, 14510 | 9 | 1 |
| 14513, 14514 | 1 | 1 |
| 14519, 14520 | 1 | 1 |
| 14597, 14598 | 12 | 4 |

| | | |
|---|---|---|
| 14603, 14604 | 94 | 5 |
| 14605, 14606 | 1 | 1 |
| 14641, 14642 | 1 | 1 |
| 14653, 14654 | 2 | 1 |
| 14693, 14694 | 1 | 1 |
| 14785, 14786 | 138 | 9 |
| 14807, 14808 | 1 | 1 |
| 14927, 14928 | 1 | 1 |
| 14999, 15000 | 1 | 1 |
| 15037, 15038 | 5 | 1 |
| 15081, 15082 | 2 | 1 |
| 15143, 15144 | 3 | 2 |
| 15167, 15168 | 1 | 1 |
| 1521, 1522 | 3 | 1 |
| 15221, 15222 | 4 | 1 |
| 15249, 15250 | 6 | 1 |
| 15309, 15310 | 1 | 1 |
| 15313, 15314 | 1 | 1 |
| 15347, 15348 | 1 | 1 |
| 15349, 15350 | 2 | 1 |
| 15353, 15354 | 1 | 1 |
| 1537, 1538 | 1 | 1 |
| 15463, 15464 | 1 | 1 |
| 15493, 15494 | 5 | 2 |
| 15497, 15498 | 1 | 1 |
| 15505, 15506 | 20 | 3 |
| 15515, 15516 | 1 | 1 |
| 15541, 15542 | 2 | 1 |
| 15543, 15544 | 1 | 1 |
| 15627, 15628 | 6 | 2 |
| 15643, 15644 | 1 | 1 |
| 15649, 15650 | 1 | 1 |
| 15687, 15688 | 1 | 1 |
| 15721, 15722 | 1 | 1 |
| 15869, 15870 | 1 | 1 |
| 159, 160 | 1 | 1 |
| 1593, 1594 | 1 | 1 |
| 15945, 15946 | 1 | 1 |
| 16019, 16020 | 1 | 1 |
| 16069, 16070 | 1 | 1 |
| 16091, 16092 | 1 | 1 |
| 16111, 16112 | 14 | 1 |
| 16183, 16184 | 4 | 1 |
| 16219, 16220 | 1 | 1 |
| 16239, 16240 | 15 | 1 |
| 16287, 16288 | 4 | 1 |
| 16323, 16324 | 1 | 1 |
| 16325, 16326 | 1 | 1 |
| 16343, 16344 | 2 | 1 |
| 1637, 1638 | 1 | 1 |
| 16413, 16414 | 1 | 1 |

| | | |
|---|---|---|
| 16427, 16428 | 1 | 1 |
| 16441, 16442 | 19 | 4 |
| 16447, 16448 | 4 | 2 |
| 16465, 16466 | 1 | 1 |
| 16467, 16468 | 18 | 2 |
| 16483, 16484 | 1 | 1 |
| 16495, 16496 | 1 | 1 |
| 1651, 1652 | 39 | 1 |
| 16529, 16530 | 10 | 1 |
| 16539, 16540 | 2 | 1 |
| 16567, 16568 | 6 | 1 |
| 16579, 16580 | 1 | 1 |
| 16645, 16646 | 7 | 1 |
| 1669, 1670 | 1 | 1 |
| 16753, 16754 | 65 | 8 |
| 16839, 16840 | 1 | 1 |
| 1689, 1690 | 2 | 2 |
| 16917, 16918 | 9 | 5 |
| 1701, 1702 | 2 | 2 |
| 17179, 17180 | 3 | 2 |
| 17205, 17206 | 1 | 1 |
| 17225, 17226 | 2 | 1 |
| 17349, 17350 | 3 | 1 |
| 17355, 17356 | 1 | 1 |
| 17363, 17364 | 3 | 1 |
| 17377, 17378 | 1 | 1 |
| 17393, 17394 | 4 | 2 |
| 17437, 17438 | 1 | 1 |
| 17449, 17450 | 1 | 1 |
| 1747, 1748 | 7 | 3 |
| 17541, 17542 | 91 | 3 |
| 17559, 17560 | 1 | 1 |
| 17625, 17626 | 1 | 1 |
| 17639, 17640 | 1 | 1 |
| 17653, 17654 | 1 | 1 |
| 17655, 17656 | 18 | 1 |
| 17683, 17684 | 1 | 1 |
| 177, 178 | 1 | 1 |
| 17723, 17724 | 1 | 1 |
| 17733, 17734 | 5 | 3 |
| 17803, 17804 | 66 | 7 |
| 17861, 17862 | 23 | 4 |
| 17871, 17872 | 20 | 3 |
| 17873, 17874 | 1 | 1 |
| 17887, 17888 | 2 | 2 |
| 17895, 17896 | 1 | 1 |
| 17941, 17942 | 1 | 1 |
| 1797, 1798 | 2 | 1 |
| 18031, 18032 | 1 | 1 |
| 18035, 18036 | 2 | 1 |
| 1809, 1810 | 5 | 1 |

| | | |
|---|---|---|
| 18181, 18182 | 2 | 1 |
| 18183, 18184 | 4 | 3 |
| 18197, 18198 | 5 | 3 |
| 18255, 18256 | 1 | 1 |
| 18287, 18288 | 1 | 1 |
| 18295, 18296 | 1 | 1 |
| 18331, 18332 | 23 | 2 |
| 18395, 18396 | 2 | 1 |
| 18435, 18436 | 1 | 1 |
| 18447, 18448 | 2 | 1 |
| 18611, 18612 | 4 | 1 |
| 18631, 18632 | 21 | 4 |
| 18645, 18646 | 1 | 1 |
| 18647, 18648 | 1 | 1 |
| 18661, 18662 | 17 | 3 |
| 18689, 18690 | 6 | 2 |
| 18749, 18750 | 5 | 1 |
| 18751, 18752 | 1 | 1 |
| 18785, 18786 | 5 | 2 |
| 18809, 18810 | 2 | 1 |
| 18883, 18884 | 6 | 4 |
| 19007, 19008 | 1 | 1 |
| 19013, 19014 | 1 | 1 |
| 1907, 1908 | 14 | 4 |
| 19093, 19094 | 2 | 1 |
| 19113, 19114 | 1 | 1 |
| 19121, 19122 | 1 | 1 |
| 19147, 19148 | 1 | 1 |
| 19159, 19160 | 4 | 1 |
| 19183, 19184 | 3 | 2 |
| 19185, 19186 | 1 | 1 |
| 1919, 1920 | 1 | 1 |
| 19193, 19194 | 2 | 2 |
| 19265, 19266 | 13 | 2 |
| 19311, 19312 | 4 | 1 |
| 19331, 19332 | 1 | 1 |
| 19357, 19358 | 4 | 2 |
| 19361, 19362 | 1 | 1 |
| 19369, 19370 | 1 | 1 |
| 1939, 1940 | 1 | 1 |
| 19415, 19416 | 1 | 1 |
| 19509, 19510 | 1 | 1 |
| 19607, 19608 | 1 | 1 |
| 19609, 19610 | 1 | 1 |
| 19625, 19626 | 1 | 1 |
| 19745, 19746 | 39 | 7 |
| 19781, 19782 | 4 | 2 |
| 1981, 1982 | 1 | 1 |
| 1983, 1984 | 8 | 1 |
| 19899, 19900 | 1 | 1 |
| 19925, 19926 | 1 | 1 |

| | | |
|---|---|---|
| 19967, 19968 | 8 | 2 |
| 20009, 20010 | 2 | 1 |
| 20011, 20012 | 1 | 1 |
| 20037, 20038 | 4 | 1 |
| 20051, 20052 | 3 | 1 |
| 20065, 20066 | 1 | 1 |
| 20085, 20086 | 1 | 1 |
| 20087, 20088 | 2 | 1 |
| 20159, 20160 | 1 | 1 |
| 20215, 20216 | 3 | 2 |
| 2027, 2028 | 1 | 1 |
| 20287, 20288 | 1 | 1 |
| 20289, 20290 | 2 | 2 |
| 20295, 20296 | 4 | 1 |
| 203, 204 | 5 | 2 |
| 20313, 20314 | 1 | 1 |
| 20319, 20320 | 26 | 7 |
| 20367, 20368 | 41 | 5 |
| 20417, 20418 | 1 | 1 |
| 2045, 2046 | 61 | 3 |
| 20479, 20480 | 54 | 5 |
| 20527, 20528 | 1 | 1 |
| 20531, 20532 | 15 | 5 |
| 2057, 2058 | 1 | 1 |
| 20581, 20582 | 1 | 1 |
| 20611, 20612 | 9 | 3 |
| 20651, 20652 | 39 | 4 |
| 20749, 20750 | 1 | 1 |
| 2077, 2078 | 3 | 1 |
| 20885, 20886 | 11 | 1 |
| 20887, 20888 | 1 | 1 |
| 20925, 20926 | 1 | 1 |
| 20963, 20964 | 1 | 1 |
| 20973, 20974 | 126 | 4 |
| 21131, 21132 | 1 | 1 |
| 2115, 2116 | 9 | 3 |
| 2117, 2118 | 2 | 1 |
| 21183, 21184 | 7 | 3 |
| 21191, 21192 | 2 | 2 |
| 21241, 21242 | 1 | 1 |
| 21245, 21246 | 3 | 1 |
| 2137, 2138 | 1 | 1 |
| 21443, 21444 | 1 | 1 |
| 21483, 21484 | 2 | 1 |
| 21531, 21532 | 3 | 3 |
| 21549, 21550 | 1 | 1 |
| 21747, 21748 | 1 | 1 |
| 21779, 21780 | 1 | 1 |
| 21801, 21802 | 1 | 1 |
| 21847, 21848 | 1 | 1 |
| 21861, 21862 | 1 | 1 |

| | | |
|---|---|---|
| 21873, 21874 | 2 | 1 |
| 21955, 21956 | 1 | 1 |
| 21971, 21972 | 2 | 1 |
| 22007, 22008 | 1 | 1 |
| 2201, 2202 | 1 | 1 |
| 221, 222 | 5 | 1 |
| 22121, 22122 | 1 | 1 |
| 22165, 22166 | 6 | 2 |
| 22203, 22204 | 1 | 1 |
| 22209, 22210 | 23 | 3 |
| 22217, 22218 | 1 | 1 |
| 22223, 22224 | 1 | 1 |
| 22237, 22238 | 9 | 2 |
| 22261, 22262 | 1 | 1 |
| 22267, 22268 | 1 | 1 |
| 2231, 2232 | 28 | 2 |
| 22363, 22364 | 1 | 1 |
| 22365, 22366 | 1 | 1 |
| 22379, 22380 | 1 | 1 |
| 22401, 22402 | 1 | 1 |
| 22403, 22404 | 2 | 1 |
| 22451, 22452 | 35 | 3 |
| 22477, 22478 | 1 | 1 |
| 22493, 22494 | 2 | 1 |
| 2253, 2254 | 2 | 1 |
| 22557, 22558 | 11 | 2 |
| 22559, 22560 | 8 | 5 |
| 22605, 22606 | 1 | 1 |
| 2283, 2284 | 1 | 1 |
| 22857, 22858 | 2 | 1 |
| 23141, 23142 | 37 | 4 |
| 23183, 23184 | 1 | 1 |
| 23235, 23236 | 1 | 1 |
| 23267, 23268 | 1 | 1 |
| 23269, 23270 | 1 | 1 |
| 233, 234 | 1 | 1 |
| 2335, 2336 | 1 | 1 |
| 2337, 2338 | 59 | 5 |
| 23381, 23382 | 1 | 1 |
| 2341, 2342 | 1 | 1 |
| 23421, 23422 | 5 | 1 |
| 23441, 23442 | 2 | 2 |
| 2347, 2348 | 1 | 1 |
| 23471, 23472 | 26 | 4 |
| 23477, 23478 | 1 | 1 |
| 235, 236 | 1 | 1 |
| 23507, 23508 | 1 | 1 |
| 23621, 23622 | 2 | 1 |
| 23651, 23652 | 104 | 9 |
| 23653, 23654 | 76 | 8 |
| 2369, 2370 | 2 | 1 |

| | | |
|---|---|---|
| 23705, 23706 | 1 | 1 |
| 23709, 23710 | 2 | 1 |
| 2373, 2374 | 1 | 1 |
| 23795, 23796 | 12 | 5 |
| 23809, 23810 | 2 | 1 |
| 23845, 23846 | 1 | 1 |
| 23847, 23848 | 1 | 1 |
| 23905, 23906 | 11 | 2 |
| 23925, 23926 | 1 | 1 |
| 23931, 23932 | 2 | 1 |
| 23961, 23962 | 12 | 6 |
| 23987, 23988 | 1 | 1 |
| 24025, 24026 | 1 | 1 |
| 24033, 24034 | 2 | 1 |
| 24041, 24042 | 10 | 3 |
| 24055, 24056 | 53 | 8 |
| 24101, 24102 | 1 | 1 |
| 24113, 24114 | 1 | 1 |
| 2415, 2416 | 1 | 1 |
| 24191, 24192 | 1 | 1 |
| 2425, 2426 | 1 | 1 |
| 24283, 24284 | 5 | 1 |
| 2433, 2434 | 2 | 2 |
| 24341, 24342 | 5 | 1 |
| 24377, 24378 | 19 | 6 |
| 24389, 24390 | 7 | 1 |
| 2439, 2440 | 15 | 5 |
| 2445, 2446 | 11 | 2 |
| 24493, 24494 | 69 | 5 |
| 24549, 24550 | 1 | 1 |
| 24579, 24580 | 1 | 1 |
| 24597, 24598 | 1 | 1 |
| 24621, 24622 | 1 | 1 |
| 24719, 24720 | 1 | 1 |
| 24767, 24768 | 2 | 1 |
| 24807, 24808 | 5 | 3 |
| 24853, 24854 | 1 | 1 |
| 24891, 24892 | 1 | 1 |
| 24899, 24900 | 1 | 1 |
| 24937, 24938 | 1 | 1 |
| 24989, 24990 | 7 | 1 |
| 25005, 25006 | 1 | 1 |
| 25063, 25064 | 1 | 1 |
| 25089, 25090 | 1 | 1 |
| 25113, 25114 | 1 | 1 |
| 25177, 25178 | 1 | 1 |
| 25183, 25184 | 2 | 1 |
| 25243, 25244 | 35 | 2 |
| 25245, 25246 | 1 | 1 |
| 25251, 25252 | 2 | 1 |
| 25279, 25280 | 1 | 1 |

| | | |
|---|---|---|
| 2529, 2530 | 1 | 1 |
| 25293, 25294 | 4 | 1 |
| 25323, 25324 | 29 | 2 |
| 25369, 25370 | 102 | 6 |
| 25373, 25374 | 1 | 1 |
| 2545, 2546 | 1 | 1 |
| 25451, 25452 | 7 | 5 |
| 25471, 25472 | 1 | 1 |
| 25479, 25480 | 1 | 1 |
| 25511, 25512 | 2 | 1 |
| 25513, 25514 | 1 | 1 |
| 2555, 2556 | 3 | 1 |
| 25633, 25634 | 14 | 2 |
| 25717, 25718 | 1 | 1 |
| 25763, 25764 | 1 | 1 |
| 25781, 25782 | 6 | 3 |
| 25799, 25800 | 2 | 2 |
| 25875, 25876 | 2 | 1 |
| 25919, 25920 | 1 | 1 |
| 25937, 25938 | 9 | 1 |
| 25947, 25948 | 5 | 1 |
| 26111, 26112 | 5 | 1 |
| 26133, 26134 | 1 | 1 |
| 26195, 26196 | 1 | 1 |
| 26213, 26214 | 69 | 9 |
| 26263, 26264 | 14 | 2 |
| 26265, 26266 | 1 | 1 |
| 26273, 26274 | 5 | 1 |
| 26317, 26318 | 1 | 1 |
| 26375, 26376 | 1 | 1 |
| 26391, 26392 | 2 | 1 |
| 26421, 26422 | 2 | 2 |
| 2645, 2646 | 14 | 6 |
| 26465, 26466 | 1 | 1 |
| 26491, 26492 | 4 | 2 |
| 26545, 26546 | 1 | 1 |
| 26581, 26582 | 1 | 1 |
| 26587, 26588 | 1 | 1 |
| 2659, 2660 | 1 | 1 |
| 26619, 26620 | 1 | 1 |
| 26671, 26672 | 1 | 1 |
| 267, 268 | 1 | 1 |
| 26705, 26706 | 1 | 1 |
| 26717, 26718 | 68 | 10 |
| 26741, 26742 | 1 | 1 |
| 26743, 26744 | 55 | 3 |
| 26745, 26746 | 63 | 1 |
| 26749, 26750 | 1 | 1 |
| 26775, 26776 | 44 | 8 |
| 2679, 2680 | 1 | 1 |
| 26791, 26792 | 1 | 1 |

| | | |
|---|---|---|
| 26841, 26842 | 2 | 2 |
| 26907, 26908 | 1 | 1 |
| 26927, 26928 | 1 | 1 |
| 26945, 26946 | 1 | 1 |
| 26971, 26972 | 51 | 7 |
| 26973, 26974 | 25 | 3 |
| 2703, 2704 | 8 | 1 |
| 27067, 27068 | 2 | 1 |
| 271, 272 | 5 | 2 |
| 27103, 27104 | 7 | 2 |
| 27153, 27154 | 1 | 1 |
| 27157, 27158 | 20 | 1 |
| 27159, 27160 | 2 | 1 |
| 27187, 27188 | 1 | 1 |
| 27229, 27230 | 1 | 1 |
| 27263, 27264 | 2 | 1 |
| 27267, 27268 | 1 | 1 |
| 27283, 27284 | 1 | 1 |
| 27383, 27384 | 1 | 1 |
| 2741, 2742 | 115 | 8 |
| 27415, 27416 | 1 | 1 |
| 27433, 27434 | 1 | 1 |
| 27461, 27462 | 2 | 2 |
| 27493, 27494 | 1 | 1 |
| 27511, 27512 | 1 | 1 |
| 27521, 27522 | 1 | 1 |
| 27631, 27632 | 1 | 1 |
| 27653, 27654 | 1 | 1 |
| 27685, 27686 | 1 | 1 |
| 27707, 27708 | 1 | 1 |
| 27793, 27794 | 1 | 1 |
| 27805, 27806 | 2 | 1 |
| 27851, 27852 | 2 | 1 |
| 27873, 27874 | 12 | 4 |
| 27923, 27924 | 5 | 1 |
| 2793, 2794 | 1 | 1 |
| 27947, 27948 | 1 | 1 |
| 27975, 27976 | 1 | 1 |
| 27981, 27982 | 2 | 1 |
| 2799, 2800 | 1 | 1 |
| 28011, 28012 | 1 | 1 |
| 28045, 28046 | 2 | 1 |
| 2809, 2810 | 1 | 1 |
| 28141, 28142 | 1 | 1 |
| 2819, 2820 | 39 | 8 |
| 2821, 2822 | 1 | 1 |
| 28243, 28244 | 2 | 1 |
| 28269, 28270 | 18 | 5 |
| 28307, 28308 | 1 | 1 |
| 28369, 28370 | 1 | 1 |
| 28403, 28404 | 1 | 1 |

| | | |
|---|---|---|
| 285, 286 | 20 | 5 |
| 28553, 28554 | 3 | 1 |
| 28715, 28716 | 1 | 1 |
| 28749, 28750 | 5 | 2 |
| 28751, 28752 | 2 | 1 |
| 28793, 28794 | 2 | 1 |
| 28835, 28836 | 2 | 1 |
| 2891, 2892 | 1 | 1 |
| 28917, 28918 | 55 | 5 |
| 28929, 28930 | 1 | 1 |
| 28949, 28950 | 4 | 2 |
| 28957, 28958 | 6 | 1 |
| 29061, 29062 | 4 | 1 |
| 29081, 29082 | 1 | 1 |
| 29091, 29092 | 4 | 2 |
| 29093, 29094 | 5 | 2 |
| 29127, 29128 | 10 | 3 |
| 29167, 29168 | 2 | 1 |
| 29169, 29170 | 2 | 1 |
| 29185, 29186 | 3 | 1 |
| 29247, 29248 | 2 | 1 |
| 29257, 29258 | 15 | 5 |
| 29259, 29260 | 1 | 1 |
| 29275, 29276 | 1 | 1 |
| 29321, 29322 | 2 | 1 |
| 29363, 29364 | 1 | 1 |
| 29383, 29384 | 1 | 1 |
| 29387, 29388 | 2 | 1 |
| 29419, 29420 | 1 | 1 |
| 29583, 29584 | 1 | 1 |
| 2961, 2962 | 1 | 1 |
| 29629, 29630 | 1 | 1 |
| 29771, 29772 | 2 | 2 |
| 29779, 29780 | 5 | 1 |
| 29789, 29790 | 1 | 1 |
| 29815, 29816 | 1 | 1 |
| 29821, 29822 | 2 | 1 |
| 29961, 29962 | 1 | 1 |
| 29975, 29976 | 8 | 2 |
| 29985, 29986 | 1 | 1 |
| 30067, 30068 | 1 | 1 |
| 30109, 30110 | 1 | 1 |
| 30155, 30156 | 11 | 1 |
| 3023, 3024 | 2 | 2 |
| 30239, 30240 | 1 | 1 |
| 30275, 30276 | 1 | 1 |
| 3031, 3032 | 3 | 2 |
| 30339, 30340 | 2 | 1 |
| 3045, 3046 | 1 | 1 |
| 30563, 30564 | 41 | 2 |
| 30653, 30654 | 1 | 1 |

| | | |
|---|---|---|
| 30765, 30766 | 1 | 1 |
| 30787, 30788 | 2 | 1 |
| 30811, 30812 | 6 | 2 |
| 30817, 30818 | 1 | 1 |
| 30841, 30842 | 1 | 1 |
| 30843, 30844 | 1 | 1 |
| 30853, 30854 | 47 | 6 |
| 30911, 30912 | 1 | 1 |
| 30945, 30946 | 1 | 1 |
| 30993, 30994 | 1 | 1 |
| 31007, 31008 | 1 | 1 |
| 31043, 31044 | 2 | 1 |
| 31047, 31048 | 2 | 1 |
| 3113, 3114 | 9 | 3 |
| 31133, 31134 | 1 | 1 |
| 31151, 31152 | 1 | 1 |
| 31175, 31176 | 1 | 1 |
| 31177, 31178 | 1 | 1 |
| 31211, 31212 | 13 | 1 |
| 31303, 31304 | 6 | 3 |
| 31395, 31396 | 1 | 1 |
| 31431, 31432 | 1 | 1 |
| 31473, 31474 | 1 | 1 |
| 31557, 31558 | 4 | 1 |
| 31605, 31606 | 1 | 1 |
| 31609, 31610 | 1 | 1 |
| 31633, 31634 | 1 | 1 |
| 31725, 31726 | 4 | 2 |
| 31789, 31790 | 10 | 3 |
| 31859, 31860 | 8 | 1 |
| 31875, 31876 | 10 | 1 |
| 31883, 31884 | 1 | 1 |
| 31893, 31894 | 21 | 7 |
| 31897, 31898 | 1 | 1 |
| 32159, 32160 | 2 | 1 |
| 32205, 32206 | 1 | 1 |
| 32261, 32262 | 1 | 1 |
| 32307, 32308 | 2 | 1 |
| 32309, 32310 | 3 | 1 |
| 32333, 32334 | 1 | 1 |
| 32353, 32354 | 1 | 1 |
| 32369, 32370 | 16 | 4 |
| 32371, 32372 | 4 | 1 |
| 32461, 32462 | 75 | 4 |
| 3247, 3248 | 1 | 1 |
| 32483, 32484 | 1 | 1 |
| 32571, 32572 | 1 | 1 |
| 32573, 32574 | 1 | 1 |
| 32605, 32606 | 1 | 1 |
| 3261, 3262 | 1 | 1 |
| 32675, 32676 | 1 | 1 |

| | | |
|---|---|---|
| 3277, 3278 | 1 | 1 |
| 3279, 3280 | 2 | 1 |
| 3281, 3282 | 1 | 1 |
| 32813, 32814 | 1 | 1 |
| 32839, 32840 | 71 | 9 |
| 32863, 32864 | 1 | 1 |
| 3287, 3288 | 1 | 1 |
| 32883, 32884 | 9 | 4 |
| 32899, 32900 | 1 | 1 |
| 3297, 3298 | 1 | 1 |
| 33019, 33020 | 1 | 1 |
| 33051, 33052 | 1 | 1 |
| 33085, 33086 | 7 | 2 |
| 33103, 33104 | 8 | 1 |
| 33143, 33144 | 4 | 1 |
| 33245, 33246 | 30 | 4 |
| 3325, 3326 | 7 | 2 |
| 33257, 33258 | 1 | 1 |
| 33303, 33304 | 1 | 1 |
| 33333, 33334 | 2 | 1 |
| 3339, 3340 | 1 | 1 |
| 33409, 33410 | 1 | 1 |
| 33421, 33422 | 6 | 3 |
| 33497, 33498 | 1 | 1 |
| 33515, 33516 | 4 | 1 |
| 33521, 33522 | 2 | 1 |
| 33529, 33530 | 21 | 5 |
| 33557, 33558 | 1 | 1 |
| 33571, 33572 | 1 | 1 |
| 33615, 33616 | 2 | 1 |
| 33677, 33678 | 1 | 1 |
| 33679, 33680 | 1 | 1 |
| 33707, 33708 | 1 | 1 |
| 33741, 33742 | 40 | 7 |
| 33743, 33744 | 20 | 2 |
| 33755, 33756 | 14 | 1 |
| 33757, 33758 | 1 | 1 |
| 33809, 33810 | 2 | 1 |
| 33853, 33854 | 1 | 1 |
| 33871, 33872 | 2 | 1 |
| 33879, 33880 | 3 | 2 |
| 33903, 33904 | 4 | 3 |
| 33925, 33926 | 9 | 1 |
| 33979, 33980 | 1 | 1 |
| 33993, 33994 | 1 | 1 |
| 33997, 33998 | 10 | 3 |
| 3405, 3406 | 1 | 1 |
| 34083, 34084 | 192 | 6 |
| 34105, 34106 | 1 | 1 |
| 34115, 34116 | 1 | 1 |
| 34163, 34164 | 1 | 1 |

| | | |
|---|---|---|
| 34197, 34198 | 1 | 1 |
| 34233, 34234 | 114 | 7 |
| 34237, 34238 | 1 | 1 |
| 34239, 34240 | 1 | 1 |
| 34305, 34306 | 1 | 1 |
| 34309, 34310 | 1 | 1 |
| 3437, 3438 | 53 | 7 |
| 34381, 34382 | 2 | 1 |
| 3439, 3440 | 13 | 3 |
| 3441, 3442 | 1 | 1 |
| 3445, 3446 | 82 | 6 |
| 34573, 34574 | 1 | 1 |
| 34581, 34582 | 1 | 1 |
| 34603, 34604 | 1 | 1 |
| 3461, 3462 | 1 | 1 |
| 3469, 3470 | 1 | 1 |
| 34775, 34776 | 4 | 1 |
| 34777, 34778 | 1 | 1 |
| 3479, 3480 | 1 | 1 |
| 34799, 34800 | 3 | 2 |
| 34833, 34834 | 2 | 1 |
| 34883, 34884 | 2 | 2 |
| 3491, 3492 | 15 | 1 |
| 34921, 34922 | 25 | 3 |
| 34947, 34948 | 1 | 1 |
| 34967, 34968 | 1 | 1 |
| 3497, 3498 | 1 | 1 |
| 34975, 34976 | 1 | 1 |
| 35009, 35010 | 1 | 1 |
| 35011, 35012 | 1 | 1 |
| 35025, 35026 | 1 | 1 |
| 35041, 35042 | 1 | 1 |
| 35123, 35124 | 1 | 1 |
| 35139, 35140 | 1 | 1 |
| 35161, 35162 | 1 | 1 |
| 35183, 35184 | 1 | 1 |
| 35291, 35292 | 1 | 1 |
| 35303, 35304 | 1 | 1 |
| 35321, 35322 | 1 | 1 |
| 35323, 35324 | 1 | 1 |
| 35465, 35466 | 1 | 1 |
| 35477, 35478 | 1 | 1 |
| 35487, 35488 | 1 | 1 |
| 35501, 35502 | 1 | 1 |
| 35503, 35504 | 4 | 2 |
| 3555, 3556 | 1 | 1 |
| 35621, 35622 | 1 | 1 |
| 35629, 35630 | 1 | 1 |
| 35651, 35652 | 4 | 1 |
| 3567, 3568 | 26 | 1 |
| 35735, 35736 | 3 | 3 |

| | | |
|---|---|---|
| 35755, 35756 | 1 | 1 |
| 35817, 35818 | 1 | 1 |
| 35905, 35906 | 2 | 2 |
| 35929, 35930 | 1 | 1 |
| 36037, 36038 | 138 | 8 |
| 36079, 36080 | 1 | 1 |
| 36155, 36156 | 1 | 1 |
| 36269, 36270 | 1 | 1 |
| 36295, 36296 | 3 | 1 |
| 36323, 36324 | 16 | 1 |
| 3633, 3634 | 2 | 2 |
| 36343, 36344 | 2 | 2 |
| 36435, 36436 | 3 | 1 |
| 36489, 36490 | 1 | 1 |
| 36497, 36498 | 1 | 1 |
| 36531, 36532 | 1 | 1 |
| 36545, 36546 | 3 | 1 |
| 36689, 36690 | 2 | 1 |
| 36789, 36790 | 1 | 1 |
| 36795, 36796 | 1 | 1 |
| 37049, 37050 | 1 | 1 |
| 37083, 37084 | 4 | 1 |
| 37089, 37090 | 1 | 1 |
| 37091, 37092 | 2 | 1 |
| 37095, 37096 | 1 | 1 |
| 37103, 37104 | 1 | 1 |
| 37195, 37196 | 1 | 1 |
| 37251, 37252 | 1 | 1 |
| 37367, 37368 | 1 | 1 |
| 37401, 37402 | 1 | 1 |
| 37427, 37428 | 1 | 1 |
| 37449, 37450 | 1 | 1 |
| 37489, 37490 | 2 | 1 |
| 37499, 37500 | 1 | 1 |
| 37515, 37516 | 1 | 1 |
| 37605, 37606 | 1 | 1 |
| 37627, 37628 | 1 | 1 |
| 37651, 37652 | 1 | 1 |
| 3769, 3770 | 2 | 2 |
| 3777, 3778 | 661 | 14 |
| 3781, 3782 | 1 | 1 |
| 37819, 37820 | 1 | 1 |
| 3783, 3784 | 3 | 1 |
| 37851, 37852 | 2 | 1 |
| 37939, 37940 | 1 | 1 |
| 38035, 38036 | 1 | 1 |
| 38055, 38056 | 3 | 3 |
| 38061, 38062 | 222 | 10 |
| 38063, 38064 | 1 | 1 |
| 38103, 38104 | 1 | 1 |
| 38109, 38110 | 1 | 1 |

| | | |
|---|---|---|
| 3813, 3814 | 1 | 1 |
| 38191, 38192 | 1 | 1 |
| 38211, 38212 | 2 | 1 |
| 38337, 38338 | 1 | 1 |
| 38357, 38358 | 1 | 1 |
| 38371, 38372 | 2 | 1 |
| 38411, 38412 | 1 | 1 |
| 38457, 38458 | 1 | 1 |
| 38461, 38462 | 2 | 1 |
| 38483, 38484 | 1 | 1 |
| 38531, 38532 | 1 | 1 |
| 38543, 38544 | 4 | 2 |
| 38559, 38560 | 1 | 1 |
| 38581, 38582 | 4 | 1 |
| 38619, 38620 | 1 | 1 |
| 38621, 38622 | 2 | 1 |
| 38637, 38638 | 51 | 5 |
| 38643, 38644 | 54 | 3 |
| 387, 388 | 5 | 2 |
| 38747, 38748 | 1 | 1 |
| 38767, 38768 | 1 | 1 |
| 38773, 38774 | 1 | 1 |
| 38777, 38778 | 1 | 1 |
| 38853, 38854 | 18 | 2 |
| 38863, 38864 | 1 | 1 |
| 38925, 38926 | 2 | 1 |
| 38929, 38930 | 1 | 1 |
| 39043, 39044 | 1 | 1 |
| 39055, 39056 | 1 | 1 |
| 39151, 39152 | 1 | 1 |
| 39159, 39160 | 14 | 2 |
| 3917, 3918 | 4 | 1 |
| 39173, 39174 | 1 | 1 |
| 39189, 39190 | 9 | 1 |
| 39199, 39200 | 3 | 1 |
| 39273, 39274 | 1 | 1 |
| 39311, 39312 | 1 | 1 |
| 39391, 39392 | 1 | 1 |
| 39397, 39398 | 1 | 1 |
| 39495, 39496 | 1 | 1 |
| 39511, 39512 | 1 | 1 |
| 39581, 39582 | 1 | 1 |
| 39599, 39600 | 1 | 1 |
| 39601, 39602 | 1 | 1 |
| 39655, 39656 | 1 | 1 |
| 39669, 39670 | 1 | 1 |
| 39729, 39730 | 29 | 7 |
| 3977, 3978 | 1 | 1 |
| 39793, 39794 | 4 | 1 |
| 39911, 39912 | 15 | 1 |
| 40025, 40026 | 1 | 1 |

| | | |
|---|---|---|
| 4013, 4014 | 275 | 8 |
| 4023, 4024 | 1 | 1 |
| 4025, 4026 | 10 | 2 |
| 40253, 40254 | 1 | 1 |
| 40319, 40320 | 40 | 2 |
| 40487, 40488 | 10 | 3 |
| 40513, 40514 | 30 | 1 |
| 40515, 40516 | 14 | 2 |
| 40521, 40522 | 1 | 1 |
| 40615, 40616 | 1 | 1 |
| 40671, 40672 | 13 | 2 |
| 40675, 40676 | 1 | 1 |
| 40717, 40718 | 1 | 1 |
| 40727, 40728 | 1 | 1 |
| 40749, 40750 | 1 | 1 |
| 40831, 40832 | 2 | 1 |
| 40839, 40840 | 19 | 2 |
| 40905, 40906 | 1 | 1 |
| 40919, 40920 | 3 | 1 |
| 40943, 40944 | 1 | 1 |
| 40949, 40950 | 1 | 1 |
| 40963, 40964 | 1 | 1 |
| 40969, 40970 | 2 | 1 |
| 41095, 41096 | 1 | 1 |
| 41103, 41104 | 1 | 1 |
| 4113, 4114 | 1 | 1 |
| 41137, 41138 | 1 | 1 |
| 41145, 41146 | 8 | 1 |
| 41165, 41166 | 1 | 1 |
| 41247, 41248 | 1 | 1 |
| 41263, 41264 | 3 | 1 |
| 41299, 41300 | 1 | 1 |
| 41303, 41304 | 1 | 1 |
| 4133, 4134 | 4 | 1 |
| 41353, 41354 | 11 | 2 |
| 41397, 41398 | 1 | 1 |
| 41423, 41424 | 1 | 1 |
| 41467, 41468 | 1 | 1 |
| 41531, 41532 | 23 | 3 |
| 41533, 41534 | 58 | 5 |
| 41539, 41540 | 1 | 1 |
| 41573, 41574 | 3 | 3 |
| 41629, 41630 | 1 | 1 |
| 41685, 41686 | 1 | 1 |
| 41743, 41744 | 1 | 1 |
| 41799, 41800 | 1 | 1 |
| 41815, 41816 | 1 | 1 |
| 41819, 41820 | 2 | 1 |
| 41893, 41894 | 2 | 1 |
| 41913, 41914 | 2 | 1 |
| 41915, 41916 | 1 | 1 |

| | | |
|---|---|---|
| 41947, 41948 | 1 | 1 |
| 4197, 4198 | 6 | 3 |
| 42215, 42216 | 1 | 1 |
| 42277, 42278 | 1 | 1 |
| 42369, 42370 | 1 | 1 |
| 42419, 42420 | 1 | 1 |
| 42539, 42540 | 1 | 1 |
| 42553, 42554 | 4 | 1 |
| 42583, 42584 | 2 | 1 |
| 42641, 42642 | 1 | 1 |
| 42653, 42654 | 1 | 1 |
| 42655, 42656 | 1 | 1 |
| 42657, 42658 | 1 | 1 |
| 42659, 42660 | 2 | 1 |
| 42689, 42690 | 1 | 1 |
| 42799, 42800 | 1 | 1 |
| 42961, 42962 | 1 | 1 |
| 43095, 43096 | 2 | 2 |
| 43107, 43108 | 33 | 1 |
| 43155, 43156 | 1 | 1 |
| 43169, 43170 | 1 | 1 |
| 43235, 43236 | 1 | 1 |
| 43351, 43352 | 2 | 1 |
| 43359, 43360 | 1 | 1 |
| 43425, 43426 | 1 | 1 |
| 4343, 4344 | 49 | 7 |
| 43443, 43444 | 1 | 1 |
| 43445, 43446 | 1 | 1 |
| 43503, 43504 | 3 | 1 |
| 43551, 43552 | 1 | 1 |
| 43695, 43696 | 1 | 1 |
| 4373, 4374 | 3 | 1 |
| 43749, 43750 | 17 | 4 |
| 43769, 43770 | 12 | 2 |
| 43829, 43830 | 5 | 2 |
| 4385, 4386 | 2 | 1 |
| 43851, 43852 | 21 | 1 |
| 4387, 4388 | 18 | 1 |
| 43883, 43884 | 6 | 1 |
| 4391, 4392 | 3 | 1 |
| 44115, 44116 | 2 | 1 |
| 44129, 44130 | 3 | 1 |
| 44159, 44160 | 3 | 1 |
| 44175, 44176 | 5 | 3 |
| 44179, 44180 | 1 | 1 |
| 44209, 44210 | 3 | 1 |
| 4423, 4424 | 1 | 1 |
| 44277, 44278 | 7 | 4 |
| 44289, 44290 | 2 | 1 |
| 44295, 44296 | 1 | 1 |
| 44313, 44314 | 9 | 2 |

| | | |
|---|---|---|
| 44373, 44374 | 1 | 1 |
| 44375, 44376 | 1 | 1 |
| 44385, 44386 | 1 | 1 |
| 44403, 44404 | 1 | 1 |
| 44405, 44406 | 1 | 1 |
| 44407, 44408 | 1 | 1 |
| 44411, 44412 | 1 | 1 |
| 44447, 44448 | 1 | 1 |
| 4451, 4452 | 7 | 2 |
| 44591, 44592 | 2 | 1 |
| 44779, 44780 | 1 | 1 |
| 44801, 44802 | 1 | 1 |
| 44809, 44810 | 1 | 1 |
| 44857, 44858 | 13 | 3 |
| 44895, 44896 | 1 | 1 |
| 44907, 44908 | 2 | 1 |
| 44935, 44936 | 2 | 1 |
| 44965, 44966 | 1 | 1 |
| 45129, 45130 | 1 | 1 |
| 45263, 45264 | 1 | 1 |
| 45265, 45266 | 1 | 1 |
| 45271, 45272 | 3 | 1 |
| 45391, 45392 | 6 | 3 |
| 45435, 45436 | 4 | 1 |
| 45481, 45482 | 2 | 2 |
| 45519, 45520 | 18 | 3 |
| 45521, 45522 | 1 | 1 |
| 45583, 45584 | 7 | 1 |
| 45633, 45634 | 1 | 1 |
| 45697, 45698 | 135 | 3 |
| 45701, 45702 | 1 | 1 |
| 45705, 45706 | 5 | 1 |
| 45711, 45712 | 1 | 1 |
| 45743, 45744 | 1 | 1 |
| 45767, 45768 | 1 | 1 |
| 45781, 45782 | 1 | 1 |
| 45795, 45796 | 1 | 1 |
| 45885, 45886 | 1 | 1 |
| 45909, 45910 | 3 | 1 |
| 45931, 45932 | 1 | 1 |
| 45943, 45944 | 1 | 1 |
| 45963, 45964 | 1 | 1 |
| 46009, 46010 | 2 | 2 |
| 46017, 46018 | 1 | 1 |
| 46107, 46108 | 1 | 1 |
| 46113, 46114 | 3 | 2 |
| 46187, 46188 | 1 | 1 |
| 46229, 46230 | 3 | 1 |
| 46263, 46264 | 4 | 2 |
| 46343, 46344 | 3 | 2 |
| 46345, 46346 | 1 | 1 |

| | | |
|---|---|---|
| 46361, 46362 | 11 | 1 |
| 46401, 46402 | 2 | 1 |
| 46411, 46412 | 7 | 1 |
| 46415, 46416 | 14 | 1 |
| 46499, 46500 | 1 | 1 |
| 46509, 46510 | 10 | 5 |
| 4651, 4652 | 3 | 1 |
| 46511, 46512 | 1 | 1 |
| 46543, 46544 | 2 | 1 |
| 46637, 46638 | 3 | 2 |
| 46643, 46644 | 1 | 1 |
| 46649, 46650 | 1 | 1 |
| 46661, 46662 | 1 | 1 |
| 46679, 46680 | 1 | 1 |
| 46743, 46744 | 1 | 1 |
| 46829, 46830 | 2 | 2 |
| 46899, 46900 | 1 | 1 |
| 46903, 46904 | 1 | 1 |
| 46935, 46936 | 1 | 1 |
| 46937, 46938 | 1 | 1 |
| 46957, 46958 | 1 | 1 |
| 4703, 4704 | 1 | 1 |
| 47053, 47054 | 1 | 1 |
| 47063, 47064 | 1 | 1 |
| 47093, 47094 | 57 | 8 |
| 47103, 47104 | 1 | 1 |
| 47115, 47116 | 1 | 1 |
| 47169, 47170 | 1 | 1 |
| 47271, 47272 | 1 | 1 |
| 47277, 47278 | 1 | 1 |
| 47311, 47312 | 1 | 1 |
| 47341, 47342 | 1 | 1 |
| 47383, 47384 | 4 | 1 |
| 47481, 47482 | 1 | 1 |
| 475, 476 | 1 | 1 |
| 47571, 47572 | 1 | 1 |
| 4759, 4760 | 1 | 1 |
| 47609, 47610 | 3 | 2 |
| 47617, 47618 | 1 | 1 |
| 47631, 47632 | 1 | 1 |
| 47647, 47648 | 2 | 1 |
| 47787, 47788 | 3 | 2 |
| 47837, 47838 | 1 | 1 |
| 4789, 4790 | 1 | 1 |
| 47961, 47962 | 2 | 2 |
| 47963, 47964 | 1 | 1 |
| 48017, 48018 | 16 | 5 |
| 48041, 48042 | 1 | 1 |
| 48071, 48072 | 2 | 1 |
| 48095, 48096 | 50 | 2 |
| 48097, 48098 | 49 | 11 |

| | | |
|---|---|---|
| 48207, 48208 | 1 | 1 |
| 48225, 48226 | 2 | 1 |
| 48279, 48280 | 3 | 2 |
| 48319, 48320 | 1 | 1 |
| 48325, 48326 | 1 | 1 |
| 48369, 48370 | 1 | 1 |
| 48395, 48396 | 2 | 1 |
| 48411, 48412 | 1 | 1 |
| 48479, 48480 | 2 | 2 |
| 4849, 4850 | 3 | 1 |
| 48539, 48540 | 5 | 1 |
| 48573, 48574 | 4 | 1 |
| 48669, 48670 | 1 | 1 |
| 48751, 48752 | 1 | 1 |
| 48753, 48754 | 17 | 3 |
| 48755, 48756 | 1 | 1 |
| 48773, 48774 | 1 | 1 |
| 48821, 48822 | 9 | 4 |
| 48825, 48826 | 1 | 1 |
| 48915, 48916 | 1 | 1 |
| 48931, 48932 | 1 | 1 |
| 48959, 48960 | 1 | 1 |
| 4897, 4898 | 1 | 1 |
| 48989, 48990 | 2 | 1 |
| 49005, 49006 | 1 | 1 |
| 49157, 49158 | 9 | 4 |
| 49237, 49238 | 3 | 3 |
| 49251, 49252 | 1 | 1 |
| 49287, 49288 | 1 | 1 |
| 49305, 49306 | 1 | 1 |
| 49373, 49374 | 1 | 1 |
| 49413, 49414 | 26 | 1 |
| 49415, 49416 | 1 | 1 |
| 49527, 49528 | 1 | 1 |
| 4959, 4960 | 7 | 3 |
| 49647, 49648 | 1 | 1 |
| 49671, 49672 | 2 | 2 |
| 4973, 4974 | 24 | 4 |
| 49743, 49744 | 1 | 1 |
| 49793, 49794 | 1 | 1 |
| 49887, 49888 | 1 | 1 |
| 49945, 49946 | 2 | 1 |
| 49947, 49948 | 2 | 1 |
| 49989, 49990 | 2 | 1 |
| 5003, 5004 | 1 | 1 |
| 50101, 50102 | 1 | 1 |
| 50103, 50104 | 1 | 1 |
| 50167, 50168 | 1 | 1 |
| 50181, 50182 | 1 | 1 |
| 50251, 50252 | 1 | 1 |
| 50253, 50254 | 2 | 1 |

| | | |
|---|---|---|
| 50329, 50330 | 20 | 2 |
| 50331, 50332 | 3 | 2 |
| 50429, 50430 | 1 | 1 |
| 5045, 5046 | 7 | 2 |
| 50609, 50610 | 1 | 1 |
| 50639, 50640 | 1 | 1 |
| 50645, 50646 | 7 | 1 |
| 50659, 50660 | 1 | 1 |
| 50687, 50688 | 3 | 1 |
| 507, 508 | 1 | 1 |
| 50711, 50712 | 40 | 2 |
| 50747, 50748 | 1 | 1 |
| 50753, 50754 | 1 | 1 |
| 50791, 50792 | 1 | 1 |
| 50923, 50924 | 3 | 2 |
| 50931, 50932 | 1 | 1 |
| 50967, 50968 | 1 | 1 |
| 50977, 50978 | 1 | 1 |
| 5099, 5100 | 5 | 3 |
| 50995, 50996 | 1 | 1 |
| 50997, 50998 | 1 | 1 |
| 51013, 51014 | 1 | 1 |
| 5103, 5104 | 1 | 1 |
| 51043, 51044 | 1 | 1 |
| 51049, 51050 | 1 | 1 |
| 5105, 5106 | 1 | 1 |
| 51065, 51066 | 1 | 1 |
| 51087, 51088 | 36 | 2 |
| 51105, 51106 | 175 | 8 |
| 51175, 51176 | 1 | 1 |
| 51183, 51184 | 1 | 1 |
| 51265, 51266 | 2 | 1 |
| 51309, 51310 | 15 | 2 |
| 51341, 51342 | 2 | 2 |
| 51343, 51344 | 1 | 1 |
| 51351, 51352 | 3 | 1 |
| 51379, 51380 | 1 | 1 |
| 51387, 51388 | 1 | 1 |
| 51409, 51410 | 1 | 1 |
| 51427, 51428 | 1 | 1 |
| 51447, 51448 | 1 | 1 |
| 51485, 51486 | 37 | 6 |
| 51495, 51496 | 1 | 1 |
| 515, 516 | 2 | 1 |
| 51523, 51524 | 1 | 1 |
| 51575, 51576 | 1 | 1 |
| 51693, 51694 | 1 | 1 |
| 5171, 5172 | 2 | 2 |
| 51723, 51724 | 1 | 1 |
| 51751, 51752 | 8 | 2 |
| 51815, 51816 | 1 | 1 |

| | | |
|---|---|---|
| 51821, 51822 | 1 | 1 |
| 51903, 51904 | 6 | 3 |
| 51905, 51906 | 7 | 1 |
| 51919, 51920 | 3 | 2 |
| 51935, 51936 | 1 | 1 |
| 52011, 52012 | 1 | 1 |
| 52027, 52028 | 5 | 1 |
| 52185, 52186 | 2 | 1 |
| 52203, 52204 | 2 | 1 |
| 52205, 52206 | 52 | 6 |
| 52227, 52228 | 1 | 1 |
| 52235, 52236 | 1 | 1 |
| 52237, 52238 | 2 | 1 |
| 52315, 52316 | 12 | 3 |
| 52331, 52332 | 1 | 1 |
| 52365, 52366 | 5 | 1 |
| 52373, 52374 | 1 | 1 |
| 52421, 52422 | 2 | 2 |
| 52461, 52462 | 1 | 1 |
| 5249, 5250 | 113 | 7 |
| 52513, 52514 | 3 | 2 |
| 52569, 52570 | 2 | 1 |
| 5259, 5260 | 2 | 1 |
| 52617, 52618 | 4 | 1 |
| 52631, 52632 | 1 | 1 |
| 52683, 52684 | 4 | 1 |
| 52729, 52730 | 1 | 1 |
| 52743, 52744 | 6 | 2 |
| 52747, 52748 | 1 | 1 |
| 52893, 52894 | 1 | 1 |
| 52955, 52956 | 1 | 1 |
| 52961, 52962 | 1 | 1 |
| 53041, 53042 | 1 | 1 |
| 53053, 53054 | 1 | 1 |
| 53081, 53082 | 1 | 1 |
| 53239, 53240 | 1 | 1 |
| 5331, 5332 | 2 | 2 |
| 53423, 53424 | 1 | 1 |
| 53449, 53450 | 1 | 1 |
| 5347, 5348 | 1 | 1 |
| 53475, 53476 | 3 | 1 |
| 53477, 53478 | 12 | 2 |
| 53479, 53480 | 1 | 1 |
| 5351, 5352 | 6 | 2 |
| 53523, 53524 | 1 | 1 |
| 53559, 53560 | 4 | 1 |
| 53561, 53562 | 1 | 1 |
| 53613, 53614 | 1 | 1 |
| 5373, 5374 | 9 | 3 |
| 5375, 5376 | 12 | 4 |
| 53769, 53770 | 17 | 3 |

| | | |
|---|---|---|
| 53773, 53774 | 1 | 1 |
| 53963, 53964 | 3 | 1 |
| 53967, 53968 | 1 | 1 |
| 53977, 53978 | 36 | 3 |
| 54013, 54014 | 1 | 1 |
| 54019, 54020 | 1 | 1 |
| 54093, 54094 | 1 | 1 |
| 54161, 54162 | 1 | 1 |
| 54239, 54240 | 1 | 1 |
| 5429, 5430 | 1 | 1 |
| 5435, 5436 | 1 | 1 |
| 54393, 54394 | 1 | 1 |
| 54399, 54400 | 2 | 1 |
| 54467, 54468 | 1 | 1 |
| 54537, 54538 | 5 | 2 |
| 54645, 54646 | 1 | 1 |
| 54679, 54680 | 1 | 1 |
| 54687, 54688 | 4 | 1 |
| 54719, 54720 | 8 | 1 |
| 54721, 54722 | 2 | 2 |
| 54725, 54726 | 4 | 1 |
| 54751, 54752 | 1 | 1 |
| 54789, 54790 | 1 | 1 |
| 5479, 5480 | 2 | 2 |
| 54837, 54838 | 34 | 4 |
| 54905, 54906 | 1 | 1 |
| 54987, 54988 | 1 | 1 |
| 55003, 55004 | 1 | 1 |
| 55149, 55150 | 3 | 1 |
| 55151, 55152 | 1 | 1 |
| 55227, 55228 | 1 | 1 |
| 55259, 55260 | 1 | 1 |
| 55281, 55282 | 2 | 1 |
| 55349, 55350 | 1 | 1 |
| 5545, 5546 | 2 | 1 |
| 55507, 55508 | 1 | 1 |
| 55549, 55550 | 1 | 1 |
| 55561, 55562 | 21 | 2 |
| 5557, 5558 | 1 | 1 |
| 55585, 55586 | 5 | 2 |
| 55601, 55602 | 9 | 4 |
| 5563, 5564 | 1 | 1 |
| 55633, 55634 | 1 | 1 |
| 55673, 55674 | 2 | 1 |
| 5573, 5574 | 1 | 1 |
| 5577, 5578 | 1 | 1 |
| 55803, 55804 | 1 | 1 |
| 55823, 55824 | 1 | 1 |
| 55885, 55886 | 1 | 1 |
| 55889, 55890 | 1 | 1 |
| 55915, 55916 | 1 | 1 |

| | | |
|---|---|---|
| 55971, 55972 | 1 | 1 |
| 55973, 55974 | 2 | 2 |
| 5625, 5626 | 1 | 1 |
| 56267, 56268 | 5 | 2 |
| 5627, 5628 | 1 | 1 |
| 56299, 56300 | 1 | 1 |
| 5637, 5638 | 1 | 1 |
| 56387, 56388 | 1 | 1 |
| 56435, 56436 | 1 | 1 |
| 5645, 5646 | 1 | 1 |
| 56545, 56546 | 3 | 1 |
| 56603, 56604 | 2 | 1 |
| 56617, 56618 | 1 | 1 |
| 56693, 56694 | 1 | 1 |
| 56781, 56782 | 2 | 1 |
| 569, 570 | 1 | 1 |
| 56925, 56926 | 4 | 2 |
| 56979, 56980 | 1 | 1 |
| 56997, 56998 | 6 | 1 |
| 57015, 57016 | 1 | 1 |
| 57031, 57032 | 1 | 1 |
| 57057, 57058 | 1 | 1 |
| 57077, 57078 | 1 | 1 |
| 57115, 57116 | 6 | 3 |
| 57261, 57262 | 1 | 1 |
| 57279, 57280 | 1 | 1 |
| 57309, 57310 | 1 | 1 |
| 57343, 57344 | 1 | 1 |
| 57449, 57450 | 1 | 1 |
| 57481, 57482 | 1 | 1 |
| 57483, 57484 | 22 | 5 |
| 57485, 57486 | 3 | 1 |
| 57587, 57588 | 37 | 5 |
| 5763, 5764 | 17 | 3 |
| 57643, 57644 | 1 | 1 |
| 57703, 57704 | 13 | 3 |
| 57783, 57784 | 9 | 1 |
| 57809, 57810 | 1 | 1 |
| 57829, 57830 | 6 | 1 |
| 57905, 57906 | 2 | 1 |
| 57913, 57914 | 1 | 1 |
| 57949, 57950 | 2 | 1 |
| 5801, 5802 | 428 | 9 |
| 58013, 58014 | 1 | 1 |
| 58021, 58022 | 1 | 1 |
| 58023, 58024 | 1 | 1 |
| 581, 582 | 1 | 1 |
| 58111, 58112 | 1 | 1 |
| 58129, 58130 | 1 | 1 |
| 58193, 58194 | 1 | 1 |
| 58315, 58316 | 1 | 1 |

| | | |
|---|---|---|
| 58331, 58332 | 1 | 1 |
| 58437, 58438 | 1 | 1 |
| 58469, 58470 | 1 | 1 |
| 5847, 5848 | 1 | 1 |
| 58515, 58516 | 1 | 1 |
| 58545, 58546 | 1 | 1 |
| 58559, 58560 | 1 | 1 |
| 58577, 58578 | 2 | 2 |
| 58637, 58638 | 1 | 1 |
| 58643, 58644 | 11 | 2 |
| 58649, 58650 | 14 | 6 |
| 58687, 58688 | 1 | 1 |
| 58689, 58690 | 14 | 1 |
| 58715, 58716 | 1 | 1 |
| 58723, 58724 | 1 | 1 |
| 58735, 58736 | 3 | 1 |
| 58753, 58754 | 1 | 1 |
| 58771, 58772 | 1 | 1 |
| 58775, 58776 | 5 | 3 |
| 58847, 58848 | 1 | 1 |
| 58855, 58856 | 1 | 1 |
| 58875, 58876 | 1 | 1 |
| 58917, 58918 | 1 | 1 |
| 59119, 59120 | 1 | 1 |
| 59253, 59254 | 2 | 1 |
| 59255, 59256 | 1 | 1 |
| 59305, 59306 | 1 | 1 |
| 59359, 59360 | 1 | 1 |
| 59373, 59374 | 1 | 1 |
| 59417, 59418 | 1 | 1 |
| 59427, 59428 | 1 | 1 |
| 59433, 59434 | 1 | 1 |
| 59473, 59474 | 1 | 1 |
| 59481, 59482 | 2 | 2 |
| 5949, 5950 | 3 | 1 |
| 59491, 59492 | 1 | 1 |
| 5953, 5954 | 1 | 1 |
| 59531, 59532 | 16 | 3 |
| 59533, 59534 | 1 | 1 |
| 59599, 59600 | 4 | 4 |
| 59607, 59608 | 3 | 1 |
| 59721, 59722 | 11 | 3 |
| 59783, 59784 | 86 | 10 |
| 59789, 59790 | 5 | 2 |
| 59793, 59794 | 4 | 2 |
| 59795, 59796 | 1 | 1 |
| 59873, 59874 | 2 | 2 |
| 59895, 59896 | 7 | 4 |
| 59921, 59922 | 1 | 1 |
| 5993, 5994 | 1 | 1 |
| 5999, 6000 | 326 | 11 |

| | | |
|---|---|---|
| 60023, 60024 | 1 | 1 |
| 60029, 60030 | 1 | 1 |
| 60141, 60142 | 1 | 1 |
| 60267, 60268 | 3 | 1 |
| 6029, 6030 | 7 | 2 |
| 60295, 60296 | 1 | 1 |
| 60309, 60310 | 16 | 4 |
| 60373, 60374 | 1 | 1 |
| 60407, 60408 | 1 | 1 |
| 60411, 60412 | 1 | 1 |
| 60429, 60430 | 1 | 1 |
| 60443, 60444 | 1 | 1 |
| 60479, 60480 | 1 | 1 |
| 60543, 60544 | 4 | 1 |
| 60545, 60546 | 8 | 1 |
| 60555, 60556 | 3 | 3 |
| 60597, 60598 | 1 | 1 |
| 60795, 60796 | 1 | 1 |
| 60851, 60852 | 8 | 5 |
| 60853, 60854 | 4 | 1 |
| 60885, 60886 | 4 | 1 |
| 609, 610 | 2 | 1 |
| 60917, 60918 | 1 | 1 |
| 6093, 6094 | 3 | 2 |
| 60943, 60944 | 1 | 1 |
| 60955, 60956 | 1 | 1 |
| 60973, 60974 | 2 | 2 |
| 61003, 61004 | 4 | 1 |
| 61015, 61016 | 1 | 1 |
| 61089, 61090 | 2 | 1 |
| 61147, 61148 | 1 | 1 |
| 61285, 61286 | 1 | 1 |
| 61293, 61294 | 1 | 1 |
| 613, 614 | 1 | 1 |
| 6133, 6134 | 1 | 1 |
| 61359, 61360 | 1 | 1 |
| 61375, 61376 | 1 | 1 |
| 61377, 61378 | 1 | 1 |
| 61397, 61398 | 1 | 1 |
| 6143, 6144 | 1 | 1 |
| 61601, 61602 | 1 | 1 |
| 61765, 61766 | 1 | 1 |
| 61769, 61770 | 1 | 1 |
| 6177, 6178 | 1 | 1 |
| 6179, 6180 | 11 | 3 |
| 61873, 61874 | 2 | 1 |
| 61881, 61882 | 1 | 1 |
| 62009, 62010 | 1 | 1 |
| 62013, 62014 | 1 | 1 |
| 6213, 6214 | 1 | 1 |
| 62231, 62232 | 1 | 1 |

| | | |
|---|---|---|
| 62249, 62250 | 1 | 1 |
| 62273, 62274 | 1 | 1 |
| 62303, 62304 | 1 | 1 |
| 62333, 62334 | 1 | 1 |
| 62363, 62364 | 1 | 1 |
| 6241, 6242 | 1 | 1 |
| 6249, 6250 | 1 | 1 |
| 62507, 62508 | 2 | 1 |
| 62561, 62562 | 1 | 1 |
| 62629, 62630 | 4 | 2 |
| 62641, 62642 | 2 | 2 |
| 62745, 62746 | 1 | 1 |
| 62813, 62814 | 1 | 1 |
| 62829, 62830 | 12 | 5 |
| 62845, 62846 | 1 | 1 |
| 62905, 62906 | 2 | 2 |
| 62927, 62928 | 1 | 1 |
| 63, 64 | 389 | 2 |
| 63023, 63024 | 1 | 1 |
| 63043, 63044 | 1 | 1 |
| 63055, 63056 | 1 | 1 |
| 63125, 63126 | 1 | 1 |
| 63145, 63146 | 1 | 1 |
| 6315, 6316 | 6 | 1 |
| 63221, 63222 | 1 | 1 |
| 63257, 63258 | 1 | 1 |
| 63307, 63308 | 6 | 2 |
| 63309, 63310 | 1 | 1 |
| 63365, 63366 | 1 | 1 |
| 63375, 63376 | 1 | 1 |
| 63389, 63390 | 1 | 1 |
| 63471, 63472 | 1 | 1 |
| 63503, 63504 | 1 | 1 |
| 63545, 63546 | 49 | 3 |
| 63645, 63646 | 1 | 1 |
| 6367, 6368 | 8 | 2 |
| 63677, 63678 | 1 | 1 |
| 63699, 63700 | 1 | 1 |
| 63831, 63832 | 1 | 1 |
| 63867, 63868 | 1 | 1 |
| 63909, 63910 | 10 | 5 |
| 63927, 63928 | 3 | 1 |
| 64055, 64056 | 2 | 1 |
| 64213, 64214 | 1 | 1 |
| 64451, 64452 | 1 | 1 |
| 64459, 64460 | 7 | 4 |
| 64465, 64466 | 1 | 1 |
| 64505, 64506 | 1 | 1 |
| 64509, 64510 | 2 | 1 |
| 64527, 64528 | 1 | 1 |
| 64629, 64630 | 1 | 1 |

| | | |
|---|---|---|
| 6465, 6466 | 7 | 3 |
| 64651, 64652 | 3 | 1 |
| 64673, 64674 | 1 | 1 |
| 64745, 64746 | 1 | 1 |
| 6475, 6476 | 4 | 1 |
| 64809, 64810 | 27 | 8 |
| 65113, 65114 | 3 | 1 |
| 65121, 65122 | 18 | 3 |
| 65139, 65140 | 1 | 1 |
| 65185, 65186 | 1 | 1 |
| 65189, 65190 | 1 | 1 |
| 65285, 65286 | 1 | 1 |
| 65329, 65330 | 1 | 1 |
| 6533, 6534 | 11 | 3 |
| 65345, 65346 | 4 | 1 |
| 6535, 6536 | 2 | 2 |
| 65381, 65382 | 1 | 1 |
| 65401, 65402 | 1 | 1 |
| 65501, 65502 | 1 | 1 |
| 65509, 65510 | 3 | 2 |
| 65573, 65574 | 1 | 1 |
| 65591, 65592 | 5 | 1 |
| 65643, 65644 | 1 | 1 |
| 65689, 65690 | 4 | 1 |
| 6579, 6580 | 4 | 1 |
| 65853, 65854 | 2 | 1 |
| 6595, 6596 | 2 | 1 |
| 66131, 66132 | 10 | 1 |
| 66165, 66166 | 1 | 1 |
| 66191, 66192 | 4 | 2 |
| 66199, 66200 | 10 | 2 |
| 66205, 66206 | 7 | 2 |
| 66307, 66308 | 1 | 1 |
| 6635, 6636 | 1 | 1 |
| 66359, 66360 | 4 | 3 |
| 66361, 66362 | 8 | 5 |
| 66385, 66386 | 1 | 1 |
| 66401, 66402 | 2 | 1 |
| 66433, 66434 | 1 | 1 |
| 66453, 66454 | 20 | 7 |
| 665, 666 | 5 | 1 |
| 66529, 66530 | 7 | 2 |
| 66543, 66544 | 1 | 1 |
| 66573, 66574 | 43 | 2 |
| 66579, 66580 | 1 | 1 |
| 66591, 66592 | 2 | 2 |
| 66603, 66604 | 1 | 1 |
| 66627, 66628 | 1 | 1 |
| 66635, 66636 | 1 | 1 |
| 66637, 66638 | 1 | 1 |
| 66801, 66802 | 8 | 1 |

| | | |
|---|---|---|
| 66883, 66884 | 176 | 12 |
| 66899, 66900 | 1 | 1 |
| 66907, 66908 | 2 | 1 |
| 66939, 66940 | 1 | 1 |
| 66945, 66946 | 3 | 1 |
| 6695, 6696 | 1 | 1 |
| 66999, 67000 | 1 | 1 |
| 67041, 67042 | 9 | 2 |
| 67111, 67112 | 1 | 1 |
| 67147, 67148 | 6 | 1 |
| 67155, 67156 | 1 | 1 |
| 67187, 67188 | 1 | 1 |
| 67199, 67200 | 1 | 1 |
| 67271, 67272 | 21 | 2 |
| 67307, 67308 | 1 | 1 |
| 67331, 67332 | 1 | 1 |
| 67339, 67340 | 1 | 1 |
| 67353, 67354 | 24 | 3 |
| 67373, 67374 | 3 | 1 |
| 67471, 67472 | 6 | 2 |
| 67541, 67542 | 2 | 2 |
| 67561, 67562 | 37 | 2 |
| 67597, 67598 | 1 | 1 |
| 67671, 67672 | 6 | 1 |
| 67707, 67708 | 13 | 3 |
| 67745, 67746 | 1 | 1 |
| 67747, 67748 | 71 | 31 |
| 67755, 67756 | 1 | 1 |
| 67761, 67762 | 2 | 2 |
| 67815, 67816 | 2 | 1 |
| 67819, 67820 | 1 | 1 |
| 67963, 67964 | 4 | 3 |
| 68019, 68020 | 9 | 3 |
| 68021, 68022 | 30 | 2 |
| 68043, 68044 | 1 | 1 |
| 68199, 68200 | 1 | 1 |
| 6827, 6828 | 1 | 1 |
| 68343, 68344 | 1 | 1 |
| 68359, 68360 | 12 | 4 |
| 68391, 68392 | 1 | 1 |
| 68411, 68412 | 14 | 5 |
| 68475, 68476 | 1 | 1 |
| 685, 686 | 3 | 1 |
| 68555, 68556 | 1 | 1 |
| 68605, 68606 | 2 | 1 |
| 68717, 68718 | 1 | 1 |
| 68839, 68840 | 1 | 1 |
| 68861, 68862 | 1 | 1 |
| 689, 690 | 2 | 1 |
| 69009, 69010 | 16 | 1 |
| 6901, 6902 | 1 | 1 |

| | | |
|---|---|---|
| 69075, 69076 | 1 | 1 |
| 6909, 6910 | 1 | 1 |
| 69137, 69138 | 1 | 1 |
| 69139, 69140 | 2 | 2 |
| 69167, 69168 | 1 | 1 |
| 69175, 69176 | 1 | 1 |
| 6919, 6920 | 1 | 1 |
| 6923, 6924 | 2 | 1 |
| 69255, 69256 | 4 | 2 |
| 69275, 69276 | 27 | 2 |
| 69277, 69278 | 68 | 2 |
| 69349, 69350 | 1 | 1 |
| 69367, 69368 | 2 | 1 |
| 69369, 69370 | 1 | 1 |
| 69423, 69424 | 1 | 1 |
| 69429, 69430 | 1 | 1 |
| 69437, 69438 | 1 | 1 |
| 69439, 69440 | 2 | 1 |
| 69449, 69450 | 1 | 1 |
| 69479, 69480 | 27 | 4 |
| 69489, 69490 | 1 | 1 |
| 69547, 69548 | 1 | 1 |
| 69559, 69560 | 1 | 1 |
| 69593, 69594 | 5 | 3 |
| 69667, 69668 | 1 | 1 |
| 69703, 69704 | 1 | 1 |
| 69715, 69716 | 1 | 1 |
| 69753, 69754 | 1 | 1 |
| 69755, 69756 | 2 | 1 |
| 69773, 69774 | 5 | 3 |
| 69835, 69836 | 1 | 1 |
| 69869, 69870 | 12 | 3 |
| 69905, 69906 | 1 | 1 |
| 69919, 69920 | 1 | 1 |
| 6993, 6994 | 4 | 2 |
| 70001, 70002 | 1 | 1 |
| 7003, 7004 | 1 | 1 |
| 70031, 70032 | 1 | 1 |
| 70049, 70050 | 1 | 1 |
| 7007, 7008 | 6 | 3 |
| 70073, 70074 | 1 | 1 |
| 7013, 7014 | 1 | 1 |
| 70147, 70148 | 1 | 1 |
| 70153, 70154 | 1 | 1 |
| 7017, 7018 | 5 | 1 |
| 7021, 7022 | 10 | 1 |
| 7027, 7028 | 1 | 1 |
| 70275, 70276 | 1 | 1 |
| 70301, 70302 | 1 | 1 |
| 70347, 70348 | 1 | 1 |
| 70351, 70352 | 1 | 1 |

| | | |
|---|---|---|
| 70429, 70430 | 45 | 2 |
| 70493, 70494 | 1 | 1 |
| 70505, 70506 | 1 | 1 |
| 70611, 70612 | 6 | 1 |
| 70623, 70624 | 1 | 1 |
| 70649, 70650 | 1 | 1 |
| 70651, 70652 | 1 | 1 |
| 70657, 70658 | 1 | 1 |
| 70713, 70714 | 1 | 1 |
| 70861, 70862 | 8 | 1 |
| 70873, 70874 | 9 | 4 |
| 70875, 70876 | 26 | 7 |
| 70909, 70910 | 4 | 1 |
| 70997, 70998 | 1 | 1 |
| 71, 72 | 4 | 1 |
| 71003, 71004 | 3 | 3 |
| 71017, 71018 | 1 | 1 |
| 71023, 71024 | 3 | 1 |
| 71049, 71050 | 1 | 1 |
| 71051, 71052 | 1 | 1 |
| 71077, 71078 | 1 | 1 |
| 71117, 71118 | 1 | 1 |
| 71179, 71180 | 107 | 7 |
| 71187, 71188 | 3 | 1 |
| 71215, 71216 | 14 | 1 |
| 71289, 71290 | 4 | 1 |
| 71297, 71298 | 1 | 1 |
| 71309, 71310 | 1 | 1 |
| 71343, 71344 | 2 | 2 |
| 71371, 71372 | 1 | 1 |
| 71387, 71388 | 4 | 1 |
| 71403, 71404 | 2 | 1 |
| 7141, 7142 | 1 | 1 |
| 71417, 71418 | 1 | 1 |
| 7153, 7154 | 1 | 1 |
| 71581, 71582 | 1 | 1 |
| 71609, 71610 | 1 | 1 |
| 71615, 71616 | 2 | 2 |
| 71651, 71652 | 2 | 2 |
| 71729, 71730 | 1 | 1 |
| 71787, 71788 | 1 | 1 |
| 71875, 71876 | 1 | 1 |
| 71887, 71888 | 1 | 1 |
| 7193, 7194 | 3 | 1 |
| 71971, 71972 | 1 | 1 |
| 71975, 71976 | 1 | 1 |
| 71983, 71984 | 5 | 2 |
| 72007, 72008 | 1 | 1 |
| 72045, 72046 | 11 | 1 |
| 72051, 72052 | 2 | 1 |
| 72077, 72078 | 107 | 2 |

| | | |
|---|---|---|
| 72105, 72106 | 1 | 1 |
| 72121, 72122 | 1 | 1 |
| 72125, 72126 | 1 | 1 |
| 72185, 72186 | 1 | 1 |
| 72237, 72238 | 2 | 2 |
| 72253, 72254 | 4 | 1 |
| 72275, 72276 | 1 | 1 |
| 72303, 72304 | 1 | 1 |
| 72309, 72310 | 10 | 2 |
| 72385, 72386 | 1 | 1 |
| 72393, 72394 | 1 | 1 |
| 72451, 72452 | 1 | 1 |
| 72465, 72466 | 1 | 1 |
| 72519, 72520 | 1552 | 62 |
| 72523, 72524 | 1 | 1 |
| 72535, 72536 | 1 | 1 |
| 7255, 7256 | 2 | 1 |
| 72571, 72572 | 1 | 1 |
| 72589, 72590 | 1 | 1 |
| 72697, 72698 | 1 | 1 |
| 72741, 72742 | 1 | 1 |
| 72789, 72790 | 1 | 1 |
| 72893, 72894 | 1 | 1 |
| 72927, 72928 | 2 | 1 |
| 72981, 72982 | 1 | 1 |
| 72993, 72994 | 1 | 1 |
| 73057, 73058 | 1 | 1 |
| 73059, 73060 | 3 | 2 |
| 73077, 73078 | 1 | 1 |
| 7313, 7314 | 1 | 1 |
| 73211, 73212 | 1 | 1 |
| 73221, 73222 | 3 | 1 |
| 73267, 73268 | 1 | 1 |
| 73315, 73316 | 1 | 1 |
| 7335, 7336 | 1 | 1 |
| 73357, 73358 | 1 | 1 |
| 73361, 73362 | 2 | 2 |
| 7337, 7338 | 2 | 1 |
| 73445, 73446 | 1 | 1 |
| 7345, 7346 | 1 | 1 |
| 73491, 73492 | 1 | 1 |
| 73495, 73496 | 1 | 1 |
| 73585, 73586 | 1 | 1 |
| 7359, 7360 | 23 | 3 |
| 73685, 73686 | 4 | 1 |
| 73689, 73690 | 1 | 1 |
| 73729, 73730 | 1 | 1 |
| 7377, 7378 | 10 | 2 |
| 73841, 73842 | 1 | 1 |
| 73895, 73896 | 17 | 6 |
| 73927, 73928 | 1 | 1 |

| | | |
|---|---|---|
| 74037, 74038 | 1 | 1 |
| 74125, 74126 | 1 | 1 |
| 7413, 7414 | 2 | 2 |
| 74143, 74144 | 1 | 1 |
| 74409, 74410 | 1 | 1 |
| 7443, 7444 | 1 | 1 |
| 74471, 74472 | 1 | 1 |
| 74477, 74478 | 1 | 1 |
| 74597, 74598 | 1 | 1 |
| 74663, 74664 | 1 | 1 |
| 74733, 74734 | 1 | 1 |
| 74791, 74792 | 13 | 6 |
| 74809, 74810 | 1 | 1 |
| 74863, 74864 | 1 | 1 |
| 75095, 75096 | 1 | 1 |
| 7515, 7516 | 2 | 1 |
| 7517, 7518 | 1 | 1 |
| 75173, 75174 | 1 | 1 |
| 7523, 7524 | 1 | 1 |
| 75337, 75338 | 1 | 1 |
| 7543, 7544 | 29 | 2 |
| 75491, 75492 | 1 | 1 |
| 75537, 75538 | 1 | 1 |
| 75641, 75642 | 1 | 1 |
| 7569, 7570 | 1 | 1 |
| 75699, 75700 | 1 | 1 |
| 76007, 76008 | 1 | 1 |
| 7603, 7604 | 162 | 14 |
| 76061, 76062 | 1 | 1 |
| 76149, 76150 | 1 | 1 |
| 7631, 7632 | 1 | 1 |
| 76363, 76364 | 1 | 1 |
| 76377, 76378 | 1 | 1 |
| 76441, 76442 | 1 | 1 |
| 76525, 76526 | 1 | 1 |
| 7657, 7658 | 1 | 1 |
| 76595, 76596 | 2 | 1 |
| 76637, 76638 | 1 | 1 |
| 76683, 76684 | 1 | 1 |
| 76719, 76720 | 1 | 1 |
| 76727, 76728 | 2 | 2 |
| 76763, 76764 | 1 | 1 |
| 76931, 76932 | 1 | 1 |
| 76933, 76934 | 1 | 1 |
| 76973, 76974 | 1 | 1 |
| 7703, 7704 | 1 | 1 |
| 77071, 77072 | 1 | 1 |
| 7715, 7716 | 5 | 3 |
| 77221, 77222 | 6 | 1 |
| 7723, 7724 | 1 | 1 |
| 77247, 77248 | 6 | 3 |

| | | |
|---|---|---|
| 77283, 77284 | 1 | 1 |
| 77349, 77350 | 1 | 1 |
| 7735, 7736 | 5 | 2 |
| 77435, 77436 | 1 | 1 |
| 77437, 77438 | 1 | 1 |
| 77467, 77468 | 1 | 1 |
| 7751, 7752 | 1 | 1 |
| 7773, 7774 | 154 | 8 |
| 77791, 77792 | 5 | 3 |
| 77829, 77830 | 2 | 1 |
| 77833, 77834 | 7 | 3 |
| 77845, 77846 | 1 | 1 |
| 78025, 78026 | 1 | 1 |
| 78053, 78054 | 1 | 1 |
| 78095, 78096 | 1 | 1 |
| 7813, 7814 | 1 | 1 |
| 78147, 78148 | 2 | 2 |
| 78239, 78240 | 1 | 1 |
| 78311, 78312 | 1 | 1 |
| 78339, 78340 | 5 | 1 |
| 78499, 78500 | 1 | 1 |
| 78511, 78512 | 1 | 1 |
| 78597, 78598 | 1 | 1 |
| 78637, 78638 | 1 | 1 |
| 78647, 78648 | 1 | 1 |
| 7883, 7884 | 5 | 1 |
| 7905, 7906 | 1 | 1 |
| 7909, 7910 | 7 | 1 |
| 7927, 7928 | 6 | 1 |
| 7929, 7930 | 1 | 1 |
| 7939, 7940 | 1 | 1 |
| 795, 796 | 7 | 3 |
| 7983, 7984 | 1 | 1 |
| 7999, 8000 | 1 | 1 |
| 8039, 8040 | 152 | 4 |
| 8109, 8110 | 2 | 1 |
| 8167, 8168 | 1 | 1 |
| 8197, 8198 | 2 | 1 |
| 8203, 8204 | 1 | 1 |
| 8253, 8254 | 1 | 1 |
| 8259, 8260 | 2 | 2 |
| 8311, 8312 | 1 | 1 |
| 8451, 8452 | 24 | 1 |
| 8481, 8482 | 29 | 4 |
| 8507, 8508 | 3 | 2 |
| 8523, 8524 | 1 | 1 |
| 8585, 8586 | 1 | 1 |
| 8611, 8612 | 1 | 1 |
| 8631, 8632 | 1 | 1 |
| 8661, 8662 | 70 | 7 |
| 8669, 8670 | 49 | 5 |

| | | |
|---|---|---|
| 8683, 8684 | 1 | 1 |
| 87, 88 | 1 | 1 |
| 873, 874 | 7 | 4 |
| 8777, 8778 | 91 | 10 |
| 8801, 8802 | 4 | 1 |
| 8809, 8810 | 3 | 1 |
| 8813, 8814 | 1 | 1 |
| 8831, 8832 | 8 | 3 |
| 8841, 8842 | 7 | 2 |
| 8847, 8848 | 4 | 1 |
| 8877, 8878 | 1 | 1 |
| 89, 90 | 79 | 2 |
| 8943, 8944 | 1 | 1 |
| 895, 896 | 1 | 1 |
| 8951, 8952 | 5 | 1 |
| 8995, 8996 | 1 | 1 |
| 9007, 9008 | 1 | 1 |
| 9009, 9010 | 1 | 1 |
| 9019, 9020 | 2 | 1 |
| 9097, 9098 | 10 | 2 |
| 9107, 9108 | 3 | 2 |
| 911, 912 | 2 | 1 |
| 9161, 9162 | 1 | 1 |
| 917, 918 | 8 | 4 |
| 9251, 9252 | 25 | 7 |
| 9259, 9260 | 1 | 1 |
| 9275, 9276 | 6 | 1 |
| 9339, 9340 | 5 | 1 |
| 9341, 9342 | 1 | 1 |
| 9433, 9434 | 1 | 1 |
| 9445, 9446 | 2 | 2 |
| 9479, 9480 | 1 | 1 |
| 9507, 9508 | 1 | 1 |
| 9511, 9512 | 2 | 2 |
| 9535, 9536 | 4 | 1 |
| 9579, 9580 | 1 | 1 |
| 9585, 9586 | 2 | 1 |
| 963, 964 | 1 | 1 |
| 9645, 9646 | 1 | 1 |
| 9697, 9698 | 1 | 1 |
| 97, 98 | 5 | 1 |
| 9779, 9780 | 1 | 1 |
| 9827, 9828 | 2 | 1 |
| 9843, 9844 | 3 | 1 |
| 9851, 9852 | 1 | 1 |
| 9897, 9898 | 1 | 1 |
| 9935, 9936 | 1 | 1 |
| 9969, 9970 | 20 | 2 |
| 9975, 9976 | 1 | 1 |

EXAMPLE 2: PASC Assay for CBH Subclone Activity Screening

The following example demonstrates an exemplary protocol for confirming activity of an enzyme of the invention, or for determining if a polypeptide or peptide is within the scope of the invention. In particular, this example describes a (Phosphoric Acid Swollen Cellulose) enzyme activity assay for cellobiohydrolase (CBH) subclone activity screening, and an assay for glucose oxidase screening, noting that the invention provides polypeptides and peptides having glucose oxidase and/or cellobiohydrolase (CBH) activity, as described in Tables 1 and 3, herein. These assays can be adapted to high through-put (HTP) screening apparatus/ assays, e.g., robotic assays.

The PASC activity assay was used to confirm the enzymatic activity of exemplary enzymes of the invention, as set forth in Table 5, below; for example, the first row shows that the exemplary polypeptide of the invention SEQ ID NO:54010 (encoded, e.g., by SEQ ID NO:54009), has a glycosyl hydrolase activity.

Assay Conditions:

- 100ul reaction volume
- 50mM pH5 buffer (sodium acetate) or 50mM pH7 buffer (sodium phosphate)
- 0.75% PASC (neutral pH)
- 20ul enzyme prep soluble fraction (1:5 dilution in reaction)
- (+) control and (-) control included
- Reaction in 96 well PCR plate with foil seal
- Overnight reaction at 37C using thermocycler (or heat block)

Analysis:

- Glucose Oxidase analysis; fluorescence measured at 545/590 using a SPECTRAMAX™ (Global Medical Instrumentation, Inc., Ramsey, Minnesota) reader

Materials:

- 96 well PCR plate (used for the reaction)
- Black 384 well plate with black bottom (used for glucose oxidase detection)
- Foil seal
- 1.5% PASC stock solution
- CBH subclones samples (lysed and spun down)
- CBHI (Megazyme International Ireland Ltd.) sample (positive control)
- Vector with no insert sample (negative control)
- 500mM pH5 sodium acetate stock solution
- 500mM pH7 sodium phosphate stock solution
- Thermocycler or heat block (at 37C for reaction)
- SPECTRAMAX™ fluorescence reader (for glucose oxidase detection)

- Glucose Oxidase kit:
  1) 800mM pH7.4 phosphate buffer stock
  2) Purified B-Glucosidase
  3) 2500/250 U/ml Glucose Oxidase/Horse Radish Peroxidase cocktail
  4) 50mM Amplex red stock (light sensitive)

Protocol for PASC Reaction:

1) Add 20ul (-) control (vector no insert) to well A1 and C1. A1 will be the negative control for reactions at pH5. C1 will be the negative control for reactions at pH7.
2) Dilute 1:20 CBHI (Megazyme International Ireland Ltd.) stock (20ul stock + 380ul water)
3) Add 20ul diluted CBHI to wells A2 and C2. A2 will be the (+) control for reactions at pH5. C2 will be the (+) control for reactions at pH7.
4) Add 20ul CBH subclone samples to rows A and C (if more than 10 samples, continue adding to rowB and D). RowA will be reactions at pH5. Row C will be reactions at pH7.
5) Add 20ul autoclaved water to wells in row A and C containing samples.
6) Add 10ul 500mM pH5 buffer stock to rowA (50mM in reaction)
7) Add 10ul 500mM pH7 buffer stock to RowC (50mM in reaction)
8) Using scissors, clip the tips of 200 ul pipet tips (Rainin Instrument, Oakland CA). Pipet tips must be modified to draw up PASC stock solution.
9) Add 50ul 1.5% PASC stock to all wells containing samples. Mix well using pipet.
10) Seal PCR plate well with foil seal and place plate in thermocycler (or heat block).
11) Let PCR plate incubate at 37C overnight.

Protocol for Glucose Oxidase Analysis:

1) Remove PCR plate from 37C incubator and spin down plate for 5 minutes at 4,000 RPM, using an Eppendorf centrifuge (Eppendorf AG, Germany).
2) Using a multi-channel pipet, transfer 25ul reaction supernatant to a black 384 well detection plate. Make sure not to transfer any of the pellet to detection plate (supernatant only).
3) Make 2x glucose oxidase cocktail (volume depends on number of samples):

| Component | [Stock] | 2x Cocktail |
|---|---|---|
| Sterile dI water | n/a | qs |
| Sodium phosphate pH7.4 | 800mM | 100mM |
| B-Glucosidase | variable | 0.01U/ml |
| GO/HRP mix | 2500/250U/ml | 10/1U/ml |
| Amplex red | 50mM | 0.1mM |

- Purified β-glucosidase is produced in-house.
- Glucose Oxidase (GO) is from Sigma (#G7141-50KU) (Sigma-Aldrich, St. Louis, MO). Dissolve all 50,000 units in 5ml 50mM phosphate pH7.4 buffer.
- Horse Radish Peroxidase (HRP) is from Sigma (#P2088-5KU). Dissolve all 5,000 units in 5ml 50mM phosphate pH7.4 buffer.
- GO and HRP are then combined in equal volumes (2,500/250 GO/HRP).

- Amplex Red is from Molecular Probe (#A22177) (Invitrogen Corporation, Carlsbad CA). Dissolve 10 mg vial in 0.777ml DMSO to obtain 50 mM stock. Store at -20C protected from light.

4) Add 25ul of 2x glucose oxidase cocktail to wells in 384 well detection plate containing reaction supernatant. Mix well and avoid formation of bubbles.
5) Let detection plate incubate at room temperature for 30 minutes. Protect plate from light during incubation.
6) Read plate on SPECTRAMAX™ fluorescence plate reader at 545/590 nm.
7) Save data as a text file and store data. Open up data in excel sheet and analyze results.
8) Data analysis:
   - CBH subclones deemed "active" must have much higher 545/590 values than those of the negative control. For example, a sample with a value of 1200 would not be considered active if the negative control is 1000. Conversely, a sample with a value of 2500 would be considered active if the negative control is 1000.
   - Samples performing with similar or higher values than the positive control are deemed "highly active."
   - All active CBH subclones will further characterization on other relevant substrates, e.g. AVICEL™ (GFS Chemicals, Inc., Powell, OH) microcrystalline cellulose.
   - Interpretation of glucose oxidase data is relatively subjective, thus it is very important to have a reliable positive and negative control each time an experiment is performed.

Table 6

| SEQ ID NO: | Glycosyl Hydrolase families | Activity on PASC | Notes |
|---|---|---|---|
| 54009, 54010 | Glycosyl hydrolase family 45 | Yes | Detected by activity |
| 42389, 42390 | Glycosyl hydrolase family 9 | Not detected | Detected by proteomic analysis |
| 78693, 78694 | Cellulase, Glycosyl hydrolase family 5 | Yes | Detected by proteomic analysis |
| 71289, 71290 | Cellulase | Not tested | Detected by proteomic analysis |
| 71287, 71288 | Cellulase | In process | Detected by proteomic analysis |
| 78691, 78692 | Cellulase | Yes | Detected by activity |
| 3985, 3986 | Cellulase | Yes | Detected by activity |
| 66171, 66172 | Cellulase, Glycosyl hydrolase family 5 | Not tested | Detected by proteomic analysis |
| 78687, 78688 | Cellulase | Yes | Detected by activity |
| 78689, 78690 | Cellulase | Yes | Detected by activity |
| 78695, 78696 | Cellulase | Yes | Detected by activity |
| 78683, 78684 | Cellulase | Yes | Detected by activity |
| 78685, 78686 | Cellulase, Glycosyl hydrolase family 5 | Yes | Detected by proteomic analysis |
| 71289, 71290 | Glycosyl hydrolase family 5 | | Detected by proteomic analysis |
| 4423, 4424 | Glycosyl hydrolase family 5 | | Detected by proteomic analysis |

A number of aspects of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other aspects are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated, synthetic or recombinant nucleic acid comprising

(a) a nucleic acid sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63% 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%), sequence identity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids (polynucleotide sequences) disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, over a region of at least about 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 or more residues, or over the full length of a protein coding sequence (transcript) or gene;

wherein the nucleic acid encodes at least one polypeptide having an enzymatic activity, or encodes a polypeptide or peptide capable of generating an antibody that binds specifically to a polypeptide having a sequence comprising any of the even numbered (amino acid) SEQ ID NO:s in the sequence listing, including from SEQ ID NO:2 through SEQ ID NO:108,699;

(b) a nucleic acid sequence that hybridizes under stringent conditions to a nucleic acid (a polynucleotide sequence) comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids (polynucleotide sequences) disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699,

wherein the nucleic acid encodes a polypeptide having an activity as set forth in Table 3, or encodes a polypeptide or peptide capable of generating an antibody that binds specifically to a polypeptide having a sequence comprising any of the even numbered SEQ ID NO:s in the sequence listing, including from SEQ ID NO:2 through SEQ ID NO:108,699,

and the stringent conditions include a wash step comprising a wash in 0.2X SSC at a temperature of about 65°C for about 15 minutes,

and optionally the nucleic acid is at least about 20, 30, 40, 50, 60, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more residues in length or the full length of the gene or transcript;

(c) a nucleic acid (polynucleotide) encoding at least one polypeptide having enzyme activity, binding activity and/or structural activity, wherein the polypeptide comprises the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, etc. and all polypeptides (amino acid sequences) disclosed in the SEQ ID listing, which include all even numbered (amino acid) SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699, or enzymatically active fragments thereof;

(d) (i) the nucleic acid (polynucleotide) of (a), (b) or (c) and encoding a polypeptide having at least one conservative amino acid substitution and retaining its enzyme activity, binding activity and/or structural activity; or, (ii) the nucleic acid of (d)(i), wherein the at least one conservative amino acid substitution comprises substituting an amino acid with another amino acid of like characteristics; or, a conservative substitution comprises: replacement of an aliphatic amino acid with another aliphatic amino acid; replacement of a Serine with a Threonine or vice versa; replacement of an acidic residue with another acidic residue; replacement of a residue bearing an amide group with another residue bearing an amide group; exchange of a basic residue with another basic residue; or replacement of an aromatic residue with another aromatic residue;

(e) the nucleic acid (polynucleotide) of (a), (b), (c) or (d) encoding a polypeptide having an enzyme activity, binding activity and/or structural activity but lacking a signal sequence, a prepro domain, a dockerin domain, and/or a carbohydrate binding module (CBM),

wherein optionally the carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module;

(f) the nucleic acid (polynucleotide) of (a), (b), (c), (d) or (e) encoding a polypeptide having an enzyme activity, binding activity and/or structural activity further comprising a heterologous sequence;

(g) the nucleic acid (polynucleotide) of (f), wherein the heterologous sequence comprises, or consists of a sequence encoding: (i) a heterologous signal sequence, a heterologous carbohydrate binding module, a heterologous dockerin domain, a heterologous catalytic domain (CD), or a combination thereof; (ii) the sequence of (ii), wherein the heterologous signal sequence, carbohydrate binding module or catalytic domain (CD) is derived from a heterologous enzyme; or, (iii) a tag, an epitope, a targeting peptide, a cleavable sequence, a detectable moiety or an enzyme;

(h) the nucleic acid (polynucleotide) of (g), wherein the heterologous carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module;

(i) the nucleic acid (polynucleotide) of (g), wherein the heterologous signal sequence targets the encoded protein to a vacuole, the endoplasmic reticulum, a chloroplast or a starch granule; or

(j) a nucleic acid sequence (polynucleotide) fully (completely) complementary to (a), (b), (c), (d), (e), (f), (g), (h) or (i).

2. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the sequence identity is at least about 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63% 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699.

3. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the sequence identity is over a region of at least about 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 or more residues, or the full length of a gene or a transcript.

4. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the nucleic acid sequence comprises the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699.

5. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the nucleic acid sequence encodes a polypeptide having the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699.

6. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the sequence identities are determined by analysis with a sequence comparison algorithm or by a visual inspection.

7. The isolated, synthetic or recombinant nucleic acid of claim 6, wherein the sequence comparison algorithm is a BLAST version 2.2.2 algorithm where a filtering setting is set to blastall -p blastp -d "nr pataa" -F F, and all other options are set to default.

8. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein at least one enzymatic activity is one of the enzymatic activities described in Table 1, Table 3 and/or Table 4.

9. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the protein is a structural or binding protein.

10. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the at least one enzymatic activity comprises aldolase, alpha-galactosidase, amidase, secondary amidase, amylase, catalase, dehalogenase, endoglucanase, epoxide hydrolase, or esterase activity.

11. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the at least one enzymatic activity comprises an activity in the carotenoid pathway.

12. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the at least one enzymatic activity comprises hydrolase, protease, peptidase, glucosidase, glycosidase, intein, isomerase, laccase, lipase, or monooxygenase activity.

13. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the at least one enzymatic activity comprises nitroreductase, nitrilase, P450 enzyme, pectate lyase, phosphatase, phospholipase, phytase, polymerase or xylanase activity.

14. The isolated, synthetic or recombinant nucleic acid of claim 1, wherein the enzyme or protein is thermostable or thermotolerant.

15. The isolated, synthetic or recombinant nucleic acid of claim 14, wherein the polypeptide retains enzyme activity, binding activity and/or structural activity under conditions comprising a temperature range from about -100°C to about -80°C, about -80°C to about -40°C, about -40°C to about -20°C, about -20°C to about 0°C, about 0°C to about 37°C, about 0°C to about 5°C, about 5°C to about 15°C, about 15°C to about 25°C, about 25°C to about 37°C, about 37°C to about 45°C, about 45°C to about 55°C, about 55°C to about 70°C, about 70°C to about 75°C, about 75°C to about 85°C, about 85°C to about 90°C, about 90°C to about 95°C, about 95°C to about 100°C, about 100°C to about 105°C, about 105°C to about 110°C, about 110°C to about 120°C, or 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C or more.

16. An isolated, synthetic or recombinant nucleic acid encoding a polypeptide having an enzyme activity, binding activity and/or structural activity, wherein the nucleic acid comprises a sequence that hybridizes under stringent conditions to a nucleic acid comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ

ID NO:108,699, and the nucleic acid encodes a polypeptide having enzyme activity, binding activity and/or structural activity.

17. The isolated, synthetic or recombinant nucleic acid of claim 16, wherein the nucleic acid is at least about 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more residues in length or the full length of the gene, protein coding sequence or transcript.

18. The isolated, synthetic or recombinant nucleic acid of claim 16, wherein the stringent conditions include a wash step comprising a wash in 0.2X SSC at a temperature of about 65°C for about 15 minutes.

19. A nucleic acid probe for identifying a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity, wherein the probe comprises at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more consecutive bases of a sequence comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, wherein the probe identifies the nucleic acid by binding or hybridization.

20. The nucleic acid probe of claim 19, wherein the probe comprises an oligonucleotide comprising at least about 10 to 50, about 20 to 60, about 30 to 70, about 40 to 80, about 60 to 100, or about 50 to 150 consecutive bases.

21. A nucleic acid probe for identifying a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, wherein the probe comprises a nucleic acid comprising at least about 10 consecutive residues of a nucleic acid sequence having at least 50% sequence identity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, wherein the

sequence identities are determined by analysis with a sequence comparison algorithm or by visual inspection.

22. The nucleic acid probe of claim 21, wherein the probe comprises an oligonucleotide comprising at least about about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 or more consecutive bases.

23. An amplification primer pair for amplifying a nucleic acid encoding a polypeptide having an enzymatic activity or encoding a protein, wherein the primer pair is capable of amplifying a nucleic acid comprising the sequence of claim 1 or claim 16, or a subsequence thereof.

24. The amplification primer pair of claim 23, wherein a member of the amplification primer pair comprises an oligonucleotide comprising at least about 10 to 50 consecutive bases of the sequence, or, about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 or more consecutive bases of the sequence.

25. An amplification primer pair, wherein the amplification primer pair comprises a first member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 42, 33, 34, 35 or more residues of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, and a second member having a sequence as set forth by about the first (the 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 42, 33, 34, 35 or more residues of the complementary strand of the first member.

26. An enzyme-encoding or protein-encoding nucleic acid generated by amplification of a polynucleotide using an amplification primer pair as set forth in claim 23 or claim 25.

27. The enzyme-encoding or protein-encoding nucleic acid of claim 26, wherein the amplification is by polymerase chain reaction (PCR).

28. The enzyme-encoding or protein-encoding nucleic acid of claim 26, wherein the nucleic acid generated by amplification of a gene library.

29. The enzyme-encoding or protein-encoding nucleic acid of claim 28, wherein the gene library is an environmental library.

30. An isolated, synthetic or recombinant polypeptide having an enzymatic activity or encoding a protein encoded by a nucleic acid as set forth in claim 26.

31. A method of amplifying a nucleic acid encoding a polypeptide having an enzymatic activity or encoding a protein comprising amplification of a template nucleic acid with an amplification primer pair capable of amplifying a nucleic acid sequence as set forth in claim 1 or claim 16, or a subsequence thereof.

32. An expression cassette comprising a nucleic acid comprising the sequence of claim 1 or claim 16.

33. A vector comprising a nucleic acid comprising the sequence of claim 1 or claim 16.

34. A cloning vehicle comprising a nucleic acid comprising the sequence of claim 1 or claim 16, wherein the cloning vehicle comprises a viral vector, a plasmid, a phage, a phagemid, a cosmid, a fosmid, a bacteriophage or an artificial chromosome.

35. The cloning vehicle of claim 34, wherein the viral vector comprises an adenovirus vector, a retroviral vector or an adeno-associated viral vector.

36. A bacterial artificial chromosome (BAC), a bacteriophage P1-derived vector (PAC), a yeast artificial chromosome (YAC) or a mammalian artificial chromosome (MAC) comprising the sequence of claim 1 or claim 16.

37. A transformed cell comprising a nucleic acid comprising the sequence of claim 1 or claim 16.

38. A transformed cell comprising an expression cassette as set forth in claim 32.

39. The transformed cell of claim 38, wherein the cell is a bacterial cell, a mammalian cell, a fungal cell, a yeast cell, an insect cell or a plant cell.

40. A transgenic non-human animal comprising the sequence of claim 1 or claim 16.

41. The transgenic non-human animal of claim 40, wherein the animal is a mouse, a goat, a rabbit, a sheep, a pig, a cow or a rat.

42. A transgenic plant comprising the sequence of claim 1 or claim 16.

43. The transgenic plant of claim 42, wherein the plant is a corn plant, a sorghum plant, a potato plant, a tomato plant, a wheat plant, an oilseed plant, a rapeseed plant, a soybean plant, a rice plant, a barley plant, a grass, or a tobacco plant.

44. A transgenic seed comprising the sequence of claim 1 or claim 16.

45. The transgenic seed of claim 44, wherein the seed is a corn seed, a wheat kernel, an oilseed, a rapeseed, a soybean seed, a palm kernel, a sunflower seed, a sesame seed, a rice, a barley, a peanut or a tobacco plant seed.

46. An antisense oligonucleotide comprising a nucleic acid sequence complementary to or capable of hybridizing under stringent conditions to the sequence of claim 1 or claim 16, or a subsequence thereof.

47. The antisense oligonucleotide of claim 46, wherein the antisense oligonucleotide is between about 10 to 50, about 20 to 60, about 30 to 70, about 40 to 80, or about 60 to 100 bases in length.

48. A method of inhibiting the translation of an enzyme-encoding or protein-encoding message in a cell comprising administering to the cell or expressing in the cell an antisense oligonucleotide comprising a nucleic acid sequence complementary to or capable of hybridizing under stringent conditions to the sequence of claim 1 or claim 16.

49. A double-stranded inhibitory RNA (RNAi) molecule comprising a subsequence of the sequence of claim 1 or claim 16, wherein optionally the RNAi is an siRNA or an miRNA molecule.

50. The double-stranded inhibitory RNA (RNAi) molecule of claim 49, wherein the RNAi is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length, wherein optionally the RNAi is an siRNA or an miRNA molecule.

51. A method of inhibiting the expression of an enzyme or a protein in a cell comprising administering to the cell or expressing in the cell a double-stranded inhibitory RNA (iRNA), wherein the RNA comprises a subsequence of the sequence of claim 1 or claim 16, wherein optionally the RNAi is an siRNA or an miRNA molecule.

52. An isolated, synthetic or recombinant polypeptide:

(a) comprising an amino acid sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more or complete (amino acid) sequence identity to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699, over a region of at least about 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or more residues, or over the full length of the polypeptide, wherein optionally the sequence identities are determined by analysis with a sequence comparison algorithm or by a visual inspection;

(b) encoded by a nucleic acid having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%,

70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, over a region of at least about 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or more residues, or encoded by a nucleic acid capable of hybridizing under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, or encoded by the nucleic acid (polynucleotide) of claim 1;

(c) the amino acid sequence of (a) or (b), and comprising at least one amino acid residue conservative substitution, and the polypeptide retains an enzyme activity, binding activity and/or structural activity;

(d) the amino acid sequence of (c), wherein the conservative substitution comprises replacement of an aliphatic amino acid with another aliphatic amino acid; replacement of a serine with a threonine or vice versa; replacement of an acidic residue with another acidic residue; replacement of a residue bearing an amide group with another residue bearing an amide group; exchange of a basic residue with another basic residue; or, replacement of an aromatic residue with another aromatic residue, or a combination thereof,

and optionally the aliphatic residue comprises Alanine, Valine, Leucine, Isoleucine or a synthetic equivalent thereof; the acidic residue comprises Aspartic acid, Glutamic acid or a synthetic equivalent thereof; the residue comprising an amide group comprises Aspartic acid, Glutamic acid or a synthetic equivalent thereof; the basic residue comprises Lysine, Arginine or a synthetic equivalent thereof; or, the aromatic residue comprises Phenylalanine, Tyrosine or a synthetic equivalent thereof;

(e) the polypeptide of (a), (b), (c) or (d) having an enzyme activity, binding activity and/or structural activity but lacking a signal sequence, a prepro domain, a dockerin domain, and/or a carbohydrate binding module (CBM),

wherein optionally the carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannanse binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module;

(f) the polypeptide of (a), (b), (c), (d) or (e) having an enzyme activity, binding activity and/or structural activity further comprising a heterologous sequence;

(g) the polypeptide of (f), wherein the heterologous sequence comprises, or consists of: (i) a heterologous signal sequence, a heterologous carbohydrate binding module, a heterologous dockerin domain, a heterologous catalytic domain (CD), or a combination thereof; (ii) the sequence of (ii), wherein the heterologous signal sequence, carbohydrate binding module or catalytic domain (CD) is derived from a heterologous lignocellulosic enzyme; and/or, (iii) a tag, an epitope, a targeting peptide, a cleavable sequence, a detectable moiety or an enzyme;

(h) the polypeptide of (g), wherein the heterologous carbohydrate binding module (CBM) comprises, or consists of, a xylan binding module, a cellulose binding module, a lignin binding module, a xylose binding module, a mannan binding module, a xyloglucan-specific module and/or a arabinofuranosidase binding module; or

(i) polypeptide of (g), wherein the heterologous signal sequence targets the encoded protein to a vacuole, the endoplasmic reticulum, a chloroplast or a starch granule.

53. The isolated, synthetic or recombinant polypeptide of claim 52, wherein the sequence identity is over a region of at least about at least about 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or is 100% sequence identity.

54. The isolated, synthetic or recombinant polypeptide of claim 52, wherein the sequence identity is over a region of at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more residues, or the full length of an enzyme.

55. The isolated, synthetic or recombinant polypeptide of claim 52, wherein the polypeptide has the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699.

56. The isolated, synthetic or recombinant polypeptide of claim 52, wherein the polypeptide has an enzymatic activity or encodes a protein.

57. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzymatic activity comprises aldolase, alpha-galactosidase, amidase, secondary amidase, amylase, catalase, dehalogenase, endoglucanase, epoxide hydrolase, or esterase activity.

58. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzymatic activity comprises an activity in the carotenoid pathway.

59. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzymatic activity comprises hydrolase, glucosidase, glycosidase, intein, isomerase, laccase, lipase, or monooxygenase activity.

60. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzymatic activity comprises nitroreductase, nitrilase, P450 enzyme, pectate lyase, phosphatase, phospholipase, phytase, polymerase or xylanase activity.

61. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzyme or the protein is thermostable or thermotolerant.

62. The isolated, synthetic or recombinant polypeptide of claim 61, wherein the polypeptide retains enzymatic activity under conditions comprising a temperature range of between about 1°C to about 5°C, between about 5°C to about 15°C, between about 15°C to about 25°C, between about 25°C to about 37°C, between about 37°C to about 95°C, between about 55°C to about 85°C, between about 70°C to about 95°C, between about 70°C to about 75°C, or between about 90°C to about 95°C, or, the polypeptide

retains enzymatic activity after exposure to a temperature in the range from between about 1°C to about 5°C, between about 5°C to about 15°C, between about 15°C to about 25°C, between about 25°C to about 37°C, between about 37°C to about 95°C, between about 55°C to about 85°C, between about 70°C to about 75°C, or between about 90°C to about 95°C, or more.

63. An isolated, synthetic or recombinant polypeptide comprising a polypeptide as set forth in claim 52 and lacking a signal sequence or a prepro sequence.

64. An isolated, synthetic or recombinant polypeptide comprising a polypeptide as set forth in claim 52 and having a heterologous signal sequence or a heterologous prepro sequence.

65. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the enzymatic activity comprises a specific activity at about 37°C in the range from about 100 to about 1000 units per milligram of protein, from about 500 to about 750 units per milligram of protein, from about 500 to about 1200 units per milligram of protein, or from about 750 to about 1000 units per milligram of protein.

66. The isolated, synthetic or recombinant polypeptide of claim 61, wherein thermotolerance comprises retention of at least half of the specific activity of the enzyme at 37°C after being heated to an elevated temperature.

67. The isolated, synthetic or recombinant polypeptide of claim 61, wherein thermotolerance comprises retention of specific activity at 37°C in the range from about 500 to about 1200 units per milligram of protein after being heated to an elevated temperature.

68. The isolated, synthetic or recombinant polypeptide of claim 52, wherein the polypeptide comprises at least one glycosylation site.

69. The isolated, synthetic or recombinant polypeptide of claim 68, wherein the glycosylation is an N-linked glycosylation.

70. The isolated, synthetic or recombinant polypeptide of claim 68, wherein the polypeptide is glycosylated after being expressed in a *P. pastoris* or a *S. pombe*.

71. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the polypeptide retains an enzymatic activity under conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 or less.

72. The isolated, synthetic or recombinant polypeptide of claim 56, wherein the polypeptide retains an enzymatic activity under conditions comprising about pH 7.5, pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11.0, pH 11.5, pH 12.0, pH 12.5 or more.

73. A protein preparation comprising a polypeptide as set forth in claim 52, wherein the protein preparation comprises a liquid, a solid or a gel.

74. A heterodimer comprising a polypeptide as set forth in claim 52 and a second domain.

75. The heterodimer of claim 74, wherein the second domain is a polypeptide and the heterodimer is a fusion protein.

76. The heterodimer of claim 74, wherein the second domain is an epitope or a tag.

77. A homodimer comprising a polypeptide as set forth in claim 52.

78. An immobilized polypeptide, wherein the polypeptide comprises the sequence of claim 52, or a subsequence thereof.

79. The immobilized polypeptide of claim 78, wherein the polypeptide is immobilized on a cell, a metal, a resin, a polymer, a ceramic, a glass, a microelectrode, a graphitic particle, a bead, a gel, a plate, an array or a capillary tube.

80. An array comprising an immobilized polypeptide as set forth in claim 52.

81. An array comprising an immobilized nucleic acid as set forth in claim 1 or claim 16.

82. An isolated, synthetic or recombinant antibody that specifically binds to a polypeptide as set forth in claim 52.

83. The isolated, synthetic or recombinant antibody of claim 82, wherein the antibody is a monoclonal or a polyclonal antibody.

84. A hybridoma comprising an antibody that specifically binds to a polypeptide as set forth in claim 52.

85. A method of isolating or identifying a polypeptide with an enzymatic activity or encoding a protein comprising the steps of:

(a) providing an antibody as set forth in claim 82;

(b) providing a sample comprising polypeptides; and

(c) contacting the sample of step (b) with the antibody of step (a) under conditions wherein the antibody can specifically bind to the polypeptide, thereby isolating or identifying a polypeptide having enzymatic activity or encoding a protein.

86. A method of making an anti-enzyme or anti-protein antibody comprising administering to a non-human animal a nucleic acid as set forth in claim 1 or claim 16 or a subsequence thereof in an amount sufficient to generate a humoral immune response, thereby making an anti-enzyme or anti-protein antibody.

87. A method of making an anti-enzyme or anti-protein antibody comprising administering to a non-human animal a polypeptide as set forth in claim 52 or a subsequence thereof in an amount sufficient to generate a humoral immune response, thereby making an anti-enzyme or anti-protein antibody.

88. A method of producing a recombinant polypeptide comprising the steps of: (a) providing a nucleic acid operably linked to a promoter, wherein the nucleic acid comprises the sequence of claim 1 or claim 16; and (b) expressing the nucleic acid of step

(a) under conditions that allow expression of the polypeptide, thereby producing a recombinant polypeptide.

89. The method of claim 88, further comprising transforming a host cell with the nucleic acid of step (a) followed by expressing the nucleic acid of step (a), thereby producing a recombinant polypeptide.

90. A method for identifying a polypeptide having enzyme activity, binding activity and/or structural activity comprising the following steps:

(a) providing a polypeptide as set forth in claim 52;

(b) providing an enzyme substrate or binding substrate; and

(c) contacting the polypeptide with the substrate of step (b) and detecting a decrease in the amount of substrate or an increase in the amount of a reaction product, wherein a decrease in the amount of the substrate or an increase in the amount of the reaction product detects a polypeptide having enzyme activity, binding activity and/or structural activity.

91. A method for identifying an enzyme or protein substrate comprising the following steps:

(a) providing a polypeptide as set forth in claim 52;

(b) providing a test substrate; and

(c) contacting the polypeptide of step (a) with the test substrate of step (b) and detecting a decrease in the amount of substrate or an increase in the amount of reaction product, wherein a decrease in the amount of the substrate or an increase in the amount of a reaction product identifies the test substrate as an enzyme or protein substrate.

92. A method of determining whether a test compound specifically binds to a polypeptide comprising the following steps:

(a) expressing a nucleic acid or a vector comprising the nucleic acid under conditions permissive for translation of the nucleic acid to a polypeptide, wherein the nucleic acid has the sequence of claim 1 or claim 16;

(b) providing a test compound;

(c) contacting the polypeptide with the test compound; and

(d) determining whether the test compound of step (b) specifically binds to the polypeptide.

93. A method of determining whether a test compound specifically binds to a polypeptide comprising the following steps:

(a) providing a polypeptide as set forth in claim 52;

(b) providing a test compound;

(c) contacting the polypeptide with the test compound; and

(d) determining whether the test compound of step (b) specifically binds to the polypeptide.

94. A method for identifying a modulator of an enzyme activity, binding activity and/or structural activity comprising the following steps:

(a) providing a polypeptide as set forth in claim 56;

(b) providing a test compound;

(c) contacting the polypeptide of step (a) with the test compound of step (b) and measuring an activity of the enzyme, wherein a change in the enzyme catalytic, binding and/or structural activity measured in the presence of the test compound compared to the activity in the absence of the test compound provides a determination that the test compound modulates the enzyme activity, binding activity and/or structural activity.

95. The method of claim 94, wherein the enzyme activity is measured by providing an enzyme substrate and detecting a decrease in the amount of the substrate or an increase in the amount of a reaction product, or, an increase in the amount of the substrate or a decrease in the amount of a reaction product.

96. The method of claim 95, wherein a decrease in the amount of the substrate or an increase in the amount of the reaction product with the test compound as compared to the amount of substrate or reaction product without the test compound identifies the test compound as an activator of an enzyme activity.

97. The method of claim 95, wherein an increase in the amount of the substrate or a decrease in the amount of the reaction product with the test compound as compared to

the amount of substrate or reaction product without the test compound identifies the test compound as an inhibitor of an enzyme activity.

98. A computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a polypeptide sequence or a nucleic acid sequence, wherein the polypeptide sequence comprises sequence as set forth in claim 52, a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16.

99. The computer system of claim 98, further comprising a sequence comparison algorithm and a data storage device having at least one reference sequence stored thereon.

100. The computer system of claim 99, wherein the sequence comparison algorithm comprises a computer program that indicates polymorphisms.

101. The computer system of claim 98, further comprising an identifier that identifies one or more features in said sequence.

102. A computer readable medium having stored thereon a polypeptide sequence or a nucleic acid sequence, wherein the polypeptide sequence comprises a polypeptide as set forth in claim 52; a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16.

103. A method for identifying a feature in a sequence comprising the steps of: (a) reading the sequence using a computer program which identifies one or more features in a sequence, wherein the sequence comprises a polypeptide sequence or a nucleic acid sequence, wherein the polypeptide sequence comprises a polypeptide as set forth in claim 52; a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16; and (b) identifying one or more features in the sequence with the computer program.

104. A method for comparing a first sequence to a second sequence comprising the steps of: (a) reading the first sequence and the second sequence through use of a computer program which compares sequences, wherein the first sequence comprises a polypeptide sequence or a nucleic acid sequence, wherein the polypeptide sequence comprises a polypeptide as set forth in claim 52 or a polypeptide encoded by a nucleic

acid as set forth in claim 1 or claim 16; and (b) determining differences between the first sequence and the second sequence with the computer program.

105. The method of claim 104, wherein the step of determining differences between the first sequence and the second sequence further comprises the step of identifying polymorphisms.

106. The method of claim 104, further comprising an identifier that identifies one or more features in a sequence.

107. The method of claim 104, comprising reading the first sequence using a computer program and identifying one or more features in the sequence.

108. A method for isolating or recovering a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein from a sample comprising the steps of:

(a) providing an amplification primer pair as set forth in claim 23 or claim 25;

(b) isolating a nucleic acid from the sample or treating the sample such that nucleic acid in the sample is accessible for hybridization to the amplification primer pair; and,

(c) combining the nucleic acid of step (b) with the amplification primer pair of step (a) and amplifying nucleic acid from the sample, thereby isolating or recovering a nucleic acid encoding a polypeptide with enzyme activity, binding activity and/or structural activity or encoding a protein from a sample,

wherein optionally, the sample is an environmental sample.

109. The method of claim 108, wherein each member of the amplification primer sequence pair comprises an oligonucleotide comprising at least about 10 to 50 consecutive bases of the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, or a subsequence thereof.

110. A method for isolating or recovering a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein from a sample comprising the steps of:

(a) providing a polynucleotide probe comprising the sequence of claim 1 or claim 16, or a subsequence thereof;

(b) isolating a nucleic acid from the sample or treating the sample such that nucleic acid in the sample is accessible for hybridization to a polynucleotide probe of step (a);

(c) combining the isolated nucleic acid or the treated sample of step (b) with the polynucleotide probe of step (a); and

(d) isolating a nucleic acid that specifically hybridizes with the polynucleotide probe of step (a), thereby isolating or recovering a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein from a sample,

wherein optionally, the sample is an environmental sample.

111. The method of claim 108 or claim 110, wherein the environmental sample comprises a water sample, a liquid sample, a soil sample, an air sample or a biological sample.

112. The method of claim 108 or claim 110, wherein the sample comprises a water sample, a liquid sample, a soil sample, an air sample or a biological sample.

113. A method of generating a variant of a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein comprising the steps of:

(a) providing a template nucleic acid comprising the sequence of claim 1 or claim 16; and

(b) modifying, deleting or adding one or more nucleotides in the template sequence, or a combination thereof, to generate a variant of the template nucleic acid.

114. The method of claim 113, further comprising expressing the variant nucleic acid to generate a variant enzyme or protein polypeptide.

115. The method of claim 113, wherein the modifications, additions or deletions are introduced by a method comprising error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR) and a combination thereof.

116. The method of claim 113, wherein the modifications, additions or deletions are introduced by a method comprising recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and a combination thereof.

117. The method of claim 113, wherein the method is iteratively repeated until an enzyme or protein having an altered or different activity or an altered or different stability from that of a polypeptide encoded by the template nucleic acid is produced.

118. The method of claim 117, wherein the variant enzyme or protein polypeptide is thermotolerant, and retains some activity after being exposed to an elevated temperature.

119. The method of claim 117, wherein the variant enzyme or protein polypeptide has increased glycosylation as compared to the enzyme or protein encoded by a template nucleic acid.

120. The method of claim 117, wherein the variant enzyme or protein polypeptide has an enzyme activity, binding activity and/or structural activity under a high temperature, wherein the enzyme encoded by the template nucleic acid is not active under the high temperature.

121. The method of claim 113, wherein the method is iteratively repeated until an enzyme or protein coding sequence having an altered codon usage from that of the template nucleic acid is produced.

122. The method of claim 113, wherein the method is iteratively repeated until an enzyme or protein gene having higher or lower level of message expression or stability from that of the template nucleic acid is produced.

123. A method for modifying codons in a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein to increase its expression in a host cell, the method comprising the following steps:

(a) providing a nucleic acid encoding a polypeptide with an enzyme activity, binding activity and/or structural activity or encoding a protein comprising the sequence of claim 1 or claim 16; and,

(b) identifying a non-preferred or a less preferred codon in the nucleic acid of step (a) and replacing it with a preferred or neutrally used codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to increase its expression in a host cell.

124. A method for modifying codons in a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, the method comprising the following steps:

(a) providing a nucleic acid encoding a polypeptide with enzyme activity, binding activity and/or structural activity comprising the sequence of claim 1 or claim 16; and,

(b) identifying a codon in the nucleic acid of step (a) and replacing it with a different codon encoding the same amino acid as the replaced codon, thereby modifying codons in a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity.

125. A method for modifying codons in a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, to increase its expression in a host cell, the method comprising the following steps:

(a) providing a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, comprising the sequence of claim 1 or claim 16; and,

(b) identifying a non-preferred or a less preferred codon in the nucleic acid of step (a) and replacing it with a preferred or neutrally used codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to increase its expression in a host cell.

126. A method for modifying a codon in a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, to decrease its expression in a host cell, the method comprising the following steps:

(a) providing a nucleic acid encoding a polypeptide having enzyme activity, binding activity and/or structural activity, comprising the sequence of claim 1 or claim 16; and

(b) identifying at least one preferred codon in the nucleic acid of step (a) and replacing it with a non-preferred or less preferred codon encoding the same amino acid as the replaced codon, wherein a preferred codon is a codon over-represented in coding sequences in genes in a host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell, thereby modifying the nucleic acid to decrease its expression in a host cell.

127. The method of claim 125, wherein the host cell is a bacterial cell, a fungal cell, an insect cell, a yeast cell, a plant cell or a mammalian cell.

128. A method for producing a library of nucleic acids encoding a plurality of modified enzyme or protein active sites or substrate binding sites, wherein the modified active sites or substrate binding sites are derived from a first nucleic acid comprising a sequence encoding a first active site or a first substrate binding site the method comprising the following steps:

(a) providing a first nucleic acid encoding a first active site or first substrate binding site, wherein the first nucleic acid sequence comprises a sequence that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID

NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and all nucleic acids disclosed in the SEQ ID listing, which include all odd numbered SEQ ID NO:s from SEQ ID NO:1 through SEQ ID NO:108,699, or a subsequence thereof, and the nucleic acid encodes an enzyme or protein active site or an enzyme or protein substrate binding site;

(b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and,

(c) using the set of mutagenic oligonucleotides to generate a set of active site-encoding or substrate binding site-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of modified enzyme or protein active sites or substrate binding sites.

129. The method of claim 128, comprising mutagenizing the first nucleic acid of step (a) by a method comprising an optimized directed evolution system, Gene Site Saturation Mutagenesis (GSSM), or a synthetic ligation reassembly (SLR).

130. The method of claim 128, comprising mutagenizing the first nucleic acid of step (a) or variants by a method comprising error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, Gene Site Saturation Mutagenesis (GSSM), synthetic ligation reassembly (SLR) and a combination thereof.

131. The method of claim 128, comprising mutagenizing the first nucleic acid of step (a) or variants by a method comprising recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and a combination thereof.

132. A method for making a small molecule comprising the following steps:

(a) providing a plurality of biosynthetic enzymes capable of synthesizing or modifying a small molecule, wherein one of the enzymes comprises an enzyme encoded by a nucleic acid comprising the sequence of claim 1 or claim 16;

(b) providing a substrate for at least one of the enzymes of step (a); and

(c) reacting the substrate of step (b) with the enzymes under conditions that facilitate a plurality of biocatalytic reactions to generate a small molecule by a series of biocatalytic reactions.

133. A method for modifying a small molecule comprising the following steps:

(a) providing an enzyme, wherein the enzyme comprises a polypeptide as set forth in claim 52, or a polypeptide encoded by a nucleic acid comprising a nucleic acid sequence as set forth in claim 1 or claim 16;

(b) providing a small molecule; and

(c) reacting the enzyme of step (a) with the small molecule of step (b) under conditions that facilitate an enzymatic reaction catalyzed by the enzyme, thereby modifying a small molecule by an enzymatic reaction.

134. The method of claim 133, comprising a plurality of small molecule substrates for the enzyme of step (a), thereby generating a library of modified small molecules produced by at least one enzymatic reaction catalyzed by the enzyme.

135. The method of claim 133, further comprising a plurality of additional enzymes under conditions that facilitate a plurality of biocatalytic reactions by the enzymes to form a library of modified small molecules produced by the plurality of enzymatic reactions.

136. The method of claim 135, further comprising the step of testing the library to determine if a particular modified small molecule which exhibits a desired activity is present within the library.

137. The method of claim 136, wherein the step of testing the library further comprises the steps of systematically eliminating all but one of the biocatalytic reactions used to produce a portion of the plurality of the modified small molecules within the

library by testing the portion of the modified small molecule for the presence or absence of the particular modified small molecule with a desired activity, and identifying at least one specific biocatalytic reaction that produces the particular modified small molecule of desired activity.

138. A method for determining a functional fragment of a polypeptide having enzyme activity, binding activity and/or structural activity, comprising the steps of:

(a) providing a polypeptide having enzyme activity, binding activity and/or structural activity comprising a polypeptide as set forth in claim 52, or a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16; and

(b) deleting a plurality of amino acid residues from the sequence of step (a) and testing the remaining subsequence for an enzyme activity, binding activity and/or structural activity, thereby determining a functional fragment of an enzyme.

139. The method of claim 138, wherein the enzyme activity is measured by providing an enzyme or binding substrate and detecting a decrease in the amount of the substrate or an increase in the amount of a reaction product.

140. A method for whole cell engineering of new or modified phenotypes by using real-time metabolic flux analysis, the method comprising the following steps:

(a) making a modified cell by modifying the genetic composition of a cell, wherein the genetic composition is modified by addition to the cell of a nucleic acid comprising the sequence of claim 1 or claim 16;

(b) culturing the modified cell to generate a plurality of modified cells;

(c) measuring at least one metabolic parameter of the cell by monitoring the cell culture of step (b) in real time; and,

(d) analyzing the data of step (c) to determine if the measured parameter differs from a comparable measurement in an unmodified cell under similar conditions, thereby identifying an engineered phenotype in the cell using real-time metabolic flux analysis.

141. The method of claim 140, wherein the genetic composition of the cell is modified by a method comprising deletion of a sequence or modification of a sequence in the cell, or, knocking out the expression of a gene.

142. The method of claim 141, further comprising selecting a cell comprising a newly engineered phenotype.

143. The method of claim 142, further comprising culturing the selected cell, thereby generating a new cell strain comprising a newly engineered phenotype.

144. An isolated, synthetic or recombinant signal sequence (signal peptide) comprising, or consisting of, the sequence of residues 1 to 14, 1 to 15, 1 to 16, 1 to 17, 1 to 18, 1 to 19, 1 to 20, 1 to 21, 1 to 22, 1 to 23, 1 to 24, 1 to 25, 1 to 26, 1 to 27, 1 to 28, 1 to 28, 1 to 30, 1 to 31, 1 to 32, 1 to 33, 1 to 34, 1 to 35, 1 to 36, 1 to 37, 1 to 38, 1 to 40, 1 to 41, 1 to 42, 1 to 43, 1 to 44, 1 to 45, 1 to 46, or 1 to 47, of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, and all polypeptides disclosed in the SEQ ID listing, which include all even numbered SEQ ID NO:s from SEQ ID NO:2 through SEQ ID NO:108,699; or comprising, or consisting of the signal sequence (signal peptide) sequences set forth in Table 2.

145. A chimeric polypeptide comprising at least a first domain comprising signal peptide (SP) having the sequence of claim 144, and at least a second domain comprising a heterologous polypeptide or peptide, wherein the heterologous polypeptide or peptide is not naturally associated with the signal peptide (SP).

146. The chimeric polypeptide of claim 145, wherein the heterologous polypeptide or peptide is not an enzyme.

147. The chimeric polypeptide of claim 145, wherein the heterologous polypeptide or peptide is amino terminal to, carboxy terminal to or on both ends of the signal peptide (SP) or an enzyme catalytic domain (CD).

148. An isolated, synthetic or recombinant nucleic acid encoding a chimeric polypeptide, wherein the chimeric polypeptide comprises at least a first domain comprising signal peptide (SP) having the sequence of claim 144, and at least a second domain comprising a heterologous polypeptide or peptide, wherein the heterologous polypeptide or peptide is not naturally associated with the signal peptide (SP).

149. A method of increasing thermotolerance or thermostability of a enzyme polypeptide, the method comprising glycosylating an enzyme, wherein the polypeptide comprises at least thirty contiguous amino acids of a polypeptide as set forth in claim 52, or a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16, thereby increasing thermotolerance or thermostability of the enzyme.

150. A method for overexpressing a recombinant enzyme in a cell comprising expressing a vector comprising a nucleic acid sequence as set forth in claim 1 or claim 16, wherein overexpression is effected by use of a high activity promoter, a dicistronic vector or by gene amplification of the vector.

151. A method of making a transgenic plant comprising the following steps:

(a) introducing a heterologous nucleic acid sequence into the cell, wherein the heterologous nucleic sequence comprises the sequence of claim 1 or claim 16, thereby producing a transformed plant cell;

(b) producing a transgenic plant from the transformed cell.

152. The method as set forth in claim 151, wherein the step (a) further comprises introducing the heterologous nucleic acid sequence by electroporation or microinjection of plant cell protoplasts.

153. The method as set forth in claim 152, wherein the step (a) comprises introducing the heterologous nucleic acid sequence directly to plant tissue by DNA particle bombardment or by using an *Agrobacterium tumefaciens* host.

154. A method of expressing a heterologous nucleic acid sequence in a plant cell comprising the following steps:

(a) transforming the plant cell with a heterologous nucleic acid sequence operably linked to a promoter, wherein the heterologous nucleic sequence comprises the sequence of claim 1 or claim 16;

(b) growing the plant under conditions wherein the heterologous nucleic acids sequence is expressed in the plant cell.

155. A method for decreasing the amount of a compound in a composition comprising the following steps:

(a) providing a polypeptide having an enzyme activity or encoding a protein as set forth in claim 52, or a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16;

(b) providing a composition comprising the compound; and

(c) contacting the polypeptide of step (a) with the composition of step (b) under conditions wherein the enzyme hydrolyzes, breaks up or otherwise processes the compound in the composition.

156. The method as set forth in claim 155, wherein the composition comprises an animal food or feed.

157. A drink or beverage comprising at least one polypeptide as set forth in claim 52.

158. The drink or beverage of claim 157, wherein the drink or beverage is a food or feed supplement.

159. A food, a feed or a nutritional supplement comprising a polypeptide as set forth in claim 52.

160. A method for utilizing an enzyme or protein as a nutritional supplement in an animal diet, the method comprising:

preparing a nutritional supplement containing an enzyme or protein comprising at least thirty contiguous amino acids of a polypeptide as set forth in claim 52; and

administering the nutritional supplement to the animal.

161. The method of claim 160, wherein the animal is a human.

162. The method of claim 160, wherein the animal is a ruminant or a monogastric animal.

163. The method of claim 160, wherein the enzyme or protein is prepared by expression of a polynucleotide encoding the enzyme or protein in an organism selected from the group consisting of a bacterium, a yeast, a plant, an insect, a fungus and an animal.

164. The method of claim 163, wherein the organism is selected from the group consisting of a *S. pombe*, *S. cerevisiae*, *Pichia pastoris*, *E. coli*, *Streptomyces* sp., *Bacillus* sp. and *Lactobacillus* sp.

165. An edible enzyme delivery matrix comprising a thermostable recombinant enzyme or protein having the sequence of claim 56.

166. A method for delivering an enzyme or protein supplement to an animal, the method comprising:

preparing an edible enzyme delivery matrix in the form of pellets comprising a granulate edible carrier and thermostable recombinant enzyme or protein having the sequence of claim 56, wherein the pellets readily disperse the enzyme or protein contained therein into aqueous media, and

administering the edible enzyme delivery matrix to the animal.

167. The method of claim 166, wherein the granulate edible carrier comprises a carrier selected from the group consisting of a grain germ, a grain germ that is spent of oil, a hay, an alfalfa, a timothy, a soy hull, a sunflower seed meal and a wheat midd.

168. The method of claim 166, wherein the granulate edible carrier comprises grain germ that is spent of oil.

169. The method of claim 166, wherein the enzyme is glycosylated to provide thermostability at pelletizing conditions.

170. The method of claim 166, wherein the delivery matrix is formed by pelletizing a mixture comprising a grain germ and an enzyme or protein.

171. The method of claim 166, wherein the pelletizing conditions include application of steam.

172. The method of claim 166, wherein the pelletizing conditions comprise application of a temperature in excess of about 80°C for about 5 minutes and the enzyme retains a specific activity of at least 350 to about 900 units per milligram of enzyme.

173. An isolated, synthetic or recombinant nucleic acid comprising a sequence encoding a polypeptide having enzyme activity, binding activity and/or structural activity and a signal sequence, wherein the nucleic acid comprises the sequence of claim 1.

174. The isolated, synthetic or recombinant nucleic acid of claim 173, wherein the signal sequence is derived from another enzyme.

175. An isolated, synthetic or recombinant nucleic acid comprising a sequence encoding a polypeptide having enzyme activity, binding activity and/or structural activity, wherein the sequence does not contain a signal sequence and the nucleic acid comprises the sequence of claim 1.

176. A pharmaceutical composition comprising a polypeptide as set forth in claim 52.

177. The pharmaceutical composition of claim 176 formulated as an edible delivery agent.

178. The pharmaceutical composition of claim 177, wherein the formulation comprises a tablet, a gel, a capsule or a geltab.

179. The pharmaceutical composition of claim 177, wherein the formulation comprises a lotion, a spray or a gel.

180. A method for delivering an enzyme or protein supplement to an animal, the method comprising:

(a) providing a cell that recombinantly generates a polypeptide as set forth in claim 56, and

(b) administering the cell or the recombinantly generated polypeptide to the animal.

181. The method as claim 180, wherein the cell is a plant cell, a bacterial cell, a yeast cell, an insect cell or an animal cell.

182. The method of claim 181, wherein the cell is selected from the group consisting of a *Schizosaccharomyces* sp., *Saccharomyces* sp., *Pichia* sp., *Escherichia* sp., *Streptomyces* sp., *Bacillus* sp. and *Lactobacillus* sp.

183. The method of claim 182, wherein the cell is *Saccharomyces pombe*.

184. The method of claim 182, wherein the cell is *Saccharomyces cerevisiae*.

185. The method of claim 182, wherein the cell is *Pichia pastoris*.

186. The method of claim 182, wherein the cell is *Escherichia coli*.

187. The method of claim 182, wherein the cell is *Bacillus cereus*.

188. A composition comprising an encapsulated formulation comprising at least one polypeptide as set forth in claim 56.

189. A method for making bioethanol, biomethanol, biopropanol and/or, bio-butanol, or a biodiesel, comprising

(A) contacting a composition comprising a cellooligsaccharide, an arabinoxylan oligomer, a lignin, a lignocellulose, a xylan, a glucan, a cellulose or a fermentable sugar with the polypeptide as set forth in claim 52, or a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16;

(B) the method of (A), wherein the composition of comprises a plant, plant product or plant derivative, and optionally the plant or plant product comprises cane sugar plants or plant products, beets or sugarbeets, wheat, corn, soybeans, potato, rice or barley,

(C) the method of (A) or (B), wherein the polypeptide has activity comprising lignocellulosic enzyme and/or cellulase, endoglucanase, cellobiohydrolase, beta-glucosidase, xylanase, mannanse, β-xylosidase and/or arabinofuranosidase activity; or

(D) the method of (A), (B) or (C), wherein the plant is a monocot or dicot, or the plant is a monocot corn, sugarcane, rice, wheat, barley, switchgrass or *Miscanthus*; or the plant is a dicot oilseed crop, soy, canola, rapeseed, flax, cotton, palm oil, sugar beet, peanut, tree, poplar or lupine.

(E) the method of (A), (B), (C) or (D), further comprising processing and/or formulating the bioethanol, biomethanol, biopropanol and/or, bio-butanol, or biodiesel, as a liquid fuel, diesel and/or a gas fuel, wherein optionally the fuel comprises a biofuel, a biodiesel, and/or a synthetic fuel, or the fuel comprises bioethanol, biomethanol, biopropanol and/or, bio-butanol; and/or a gasoline-ethanol, -methanol, -butanol and/or -propanol mix.

190. A method for processing a biomass material comprising a cellooligsaccharide, an arabinoxylan oligomer, a lignin, a lignocellulose, a xylan, a glucan, a cellulose or a fermentable sugar comprising contacting the composition with the polypeptide as set forth in claim 52, or a polypeptide encoded by a nucleic acid as set forth in claim 1 or claim 16,

wherein optionally the biomass material comprising lignocellulose is derived from an agricultural crop, is a byproduct of a food or a feed production, is a lignocellulosic waste product, or is a plant residue or a waste paper or waste paper product, and optionally the polypeptide has activity comprising lignocellulosic enzyme and/or cellulase, endoglucanase, cellobiohydrolase, beta-glucosidase, xylanase, mannanse, β-xylosidase and/or arabinofuranosidase activity,

and optionally the plant residue comprise stems, leaves, hulls, husks, corn or corn cobs, corn stover, hay, straw, wood, wood chips, wood pulp, wood waste and sawdust,

and optionally the paper waste comprises discarded or used photocopy paper, computer printer paper, notebook paper, notepad paper, typewriter paper, newspapers, magazines, cardboard and paper-based packaging materials,

and optionally the processing of the biomass material generates a biodiesel, a bioalcohol, a bioethanol, biomethanol, biobutanol or biopropanol.

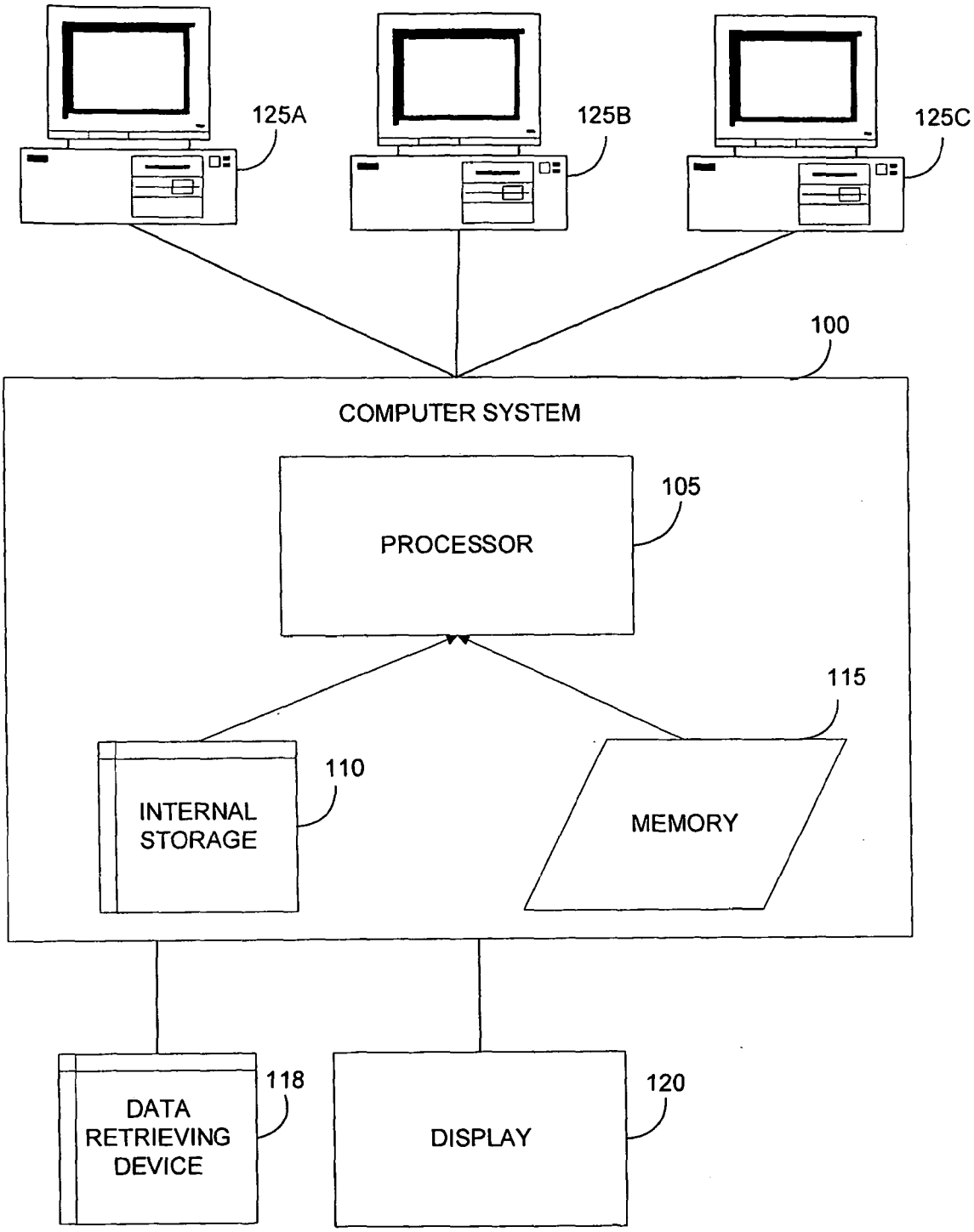

125A
125B
125C
100
COMPUTER SYSTEM
PROCESSOR
105
115
INTERNAL STORAGE
110
MEMORY
DATA RETRIEVING DEVICE
118
DISPLAY
120

FIGURE 1

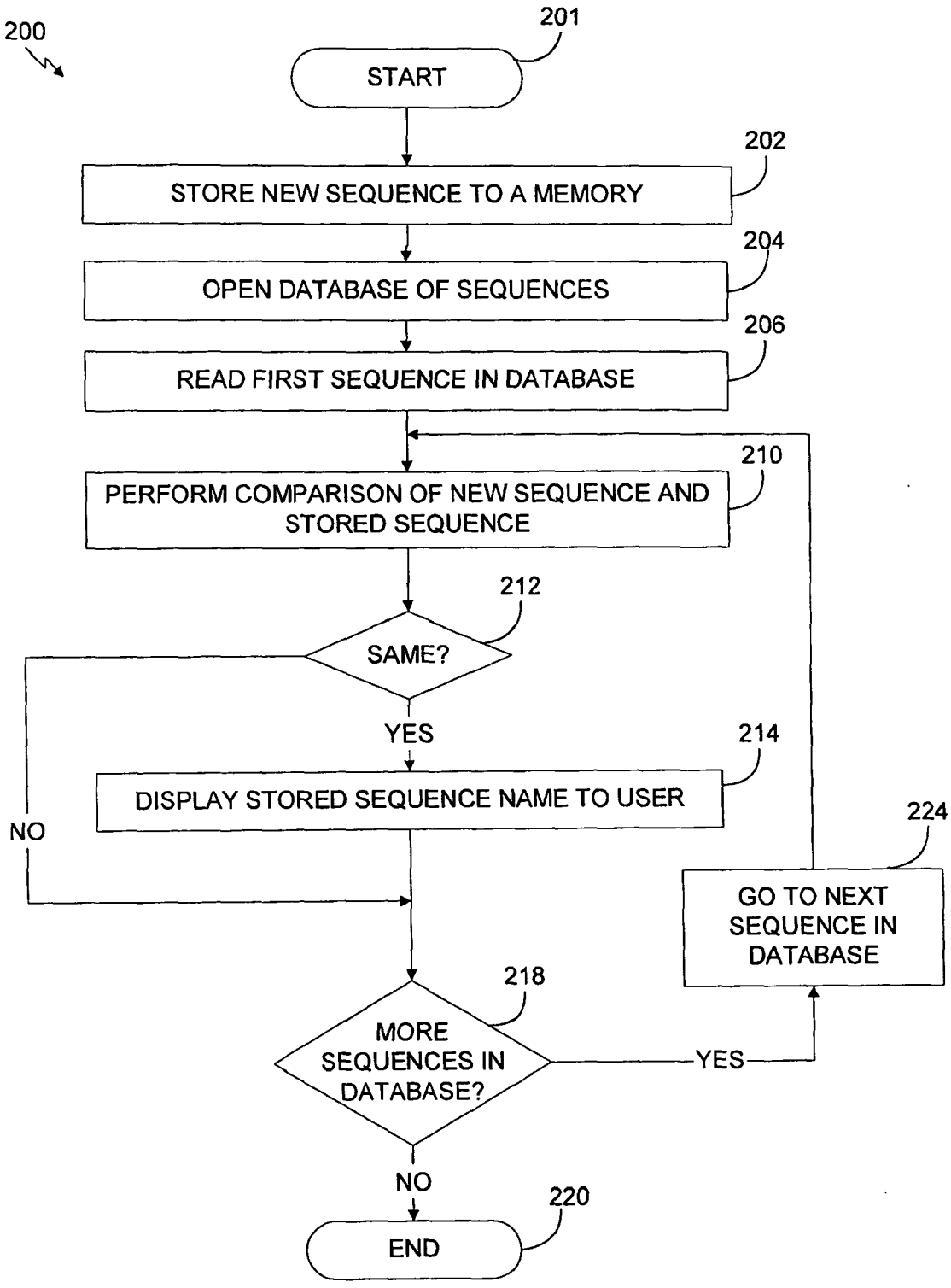

200
201
START
202
STORE NEW SEQUENCE TO A MEMORY
204
OPEN DATABASE OF SEQUENCES
206
READ FIRST SEQUENCE IN DATABASE
210
PERFORM COMPARISON OF NEW SEQUENCE AND STORED SEQUENCE
212
SAME?
YES
214
DISPLAY STORED SEQUENCE NAME TO USER
NO
224
GO TO NEXT SEQUENCE IN DATABASE
218
MORE SEQUENCES IN DATABASE?
YES
NO
220
END

FIGURE 2

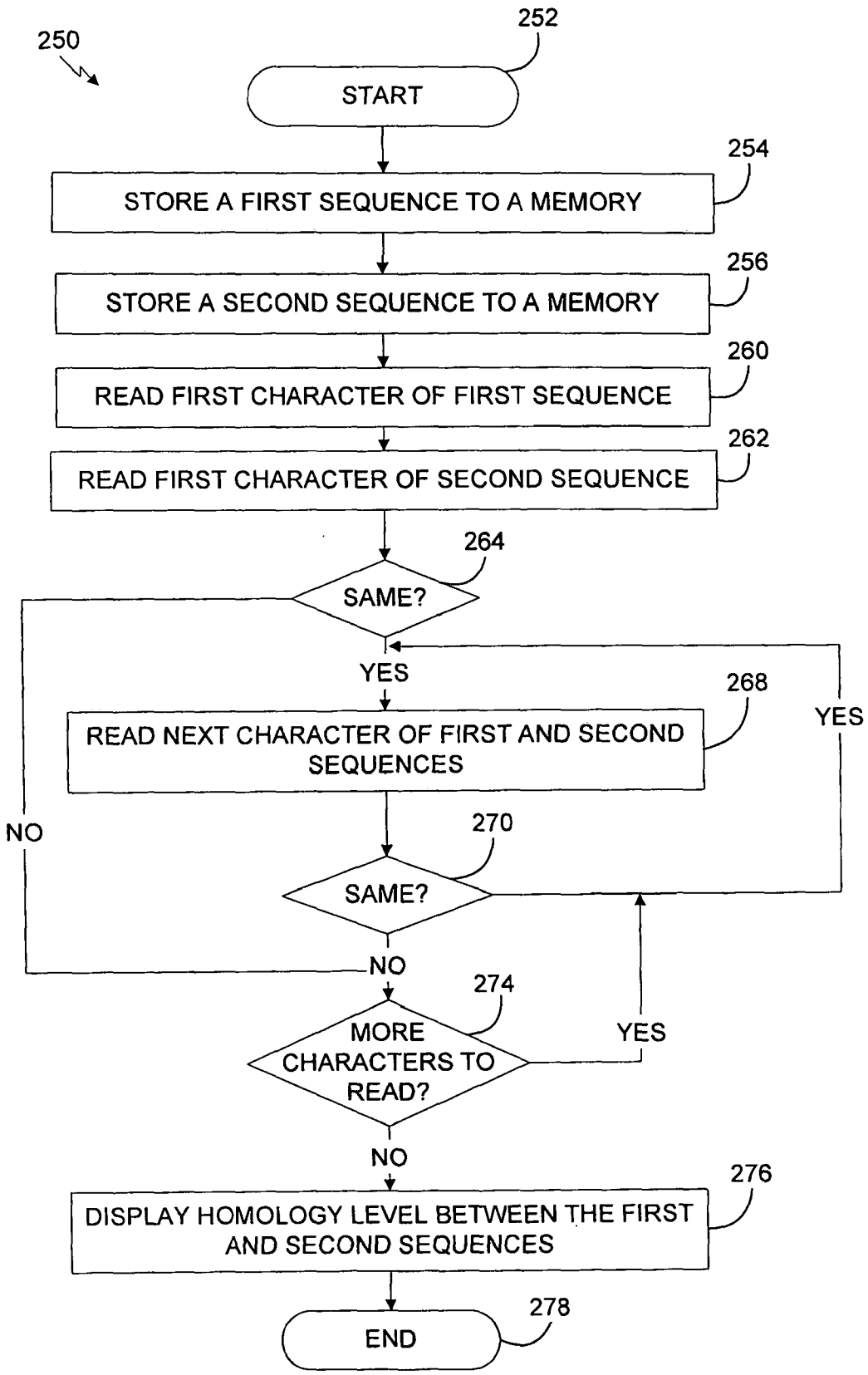

250
252
START
254
STORE A FIRST SEQUENCE TO A MEMORY
256
STORE A SECOND SEQUENCE TO A MEMORY
260
READ FIRST CHARACTER OF FIRST SEQUENCE
262
READ FIRST CHARACTER OF SECOND SEQUENCE
264
SAME?
YES
NO
268
READ NEXT CHARACTER OF FIRST AND SECOND SEQUENCES
YES
270
SAME?
NO
274
MORE CHARACTERS TO READ?
YES
NO
276
DISPLAY HOMOLOGY LEVEL BETWEEN THE FIRST AND SECOND SEQUENCES
278
END

**FIGURE 3**

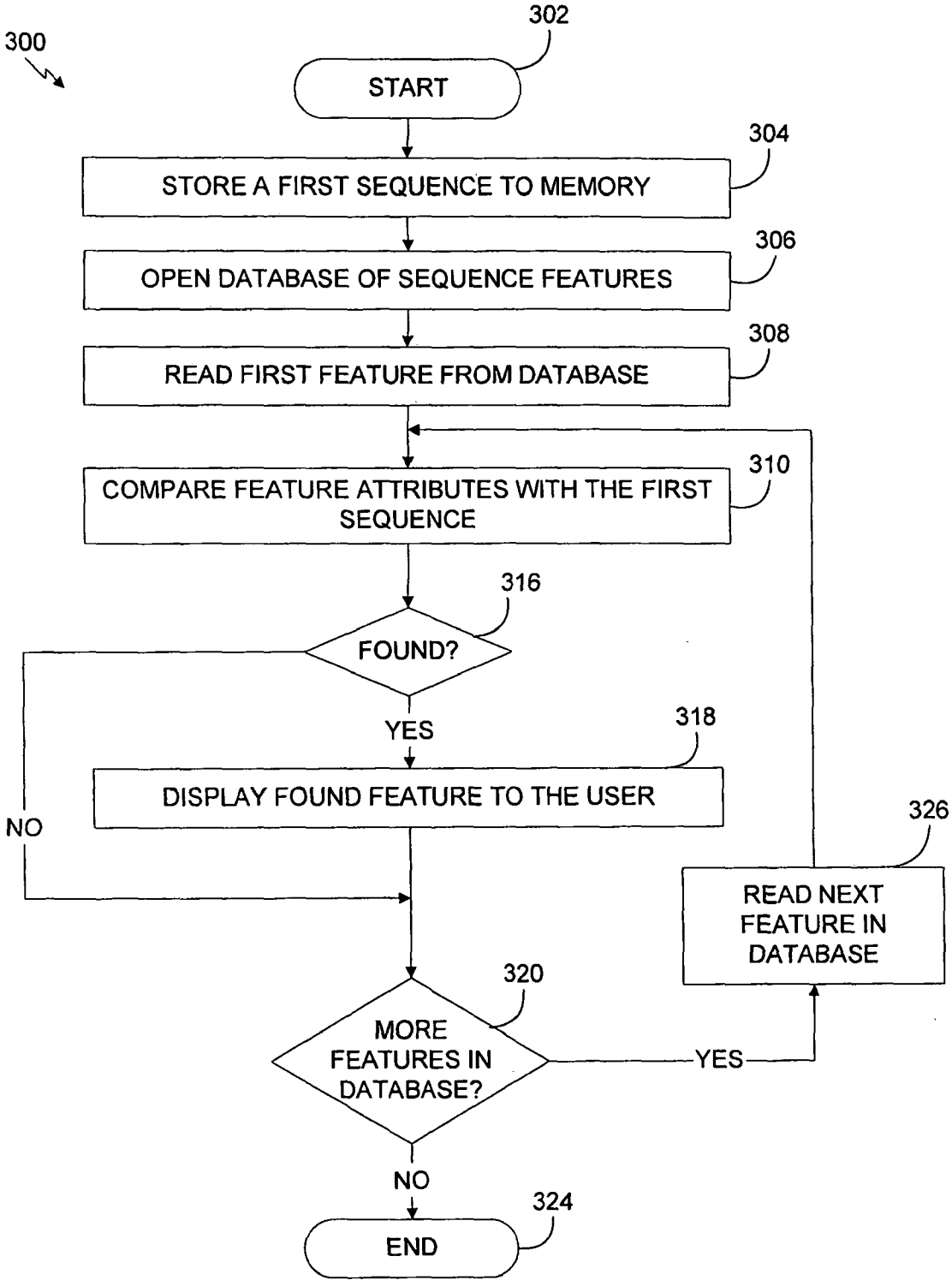

300
302
START
304
STORE A FIRST SEQUENCE TO MEMORY
306
OPEN DATABASE OF SEQUENCE FEATURES
308
READ FIRST FEATURE FROM DATABASE
310
COMPARE FEATURE ATTRIBUTES WITH THE FIRST SEQUENCE
316
FOUND?
YES
318
DISPLAY FOUND FEATURE TO THE USER
NO
326
READ NEXT FEATURE IN DATABASE
320
MORE FEATURES IN DATABASE?
YES
NO
324
END

FIGURE 4