



(51) International Patent Classification:

C12N 1/16 (2006.01) C07K 14/81 (2006.01)

C12N 15/81 (2006.01) A23J 1/18 (2006.01)

C12R 1/84 (2006.01) A23L 33/195 (2016.01)

C07K 14/395 (2006.01)

(21) International Application Number:

PCT/US2022/038095

(22) International Filing Date:

22 July 2022 (22.07.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/225,355 23 July 2021 (23.07.2021) US

63/356,944 29 June 2022 (29.06.2022) US

(71) Applicant: CLARA FOODS CO. [US/US]; 1 Tower Place, Suite 800, South San Francisco, California 94080 (US).

(72) Inventors: HURST, Logan; c/o CLARA FOODS CO., 1 Tower Place, Suite 800, South San Francisco, California 94080 (US). ZHONG, Weixi; c/o CLARA FOODS CO., 1 Tower Place, Suite 800, South San Francisco, California 94080 (US).

(74) Agent: SHARMA, Swati T.; 650 Page Mill Road, Palo Alto, California 94304 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: PROTEIN COMPOSITIONS AND METHODS OF PRODUCTION

(57) Abstract: Provided are systems and methods for production of recombinant proteins in engineered microorganisms while reducing impurities produced in the culture.

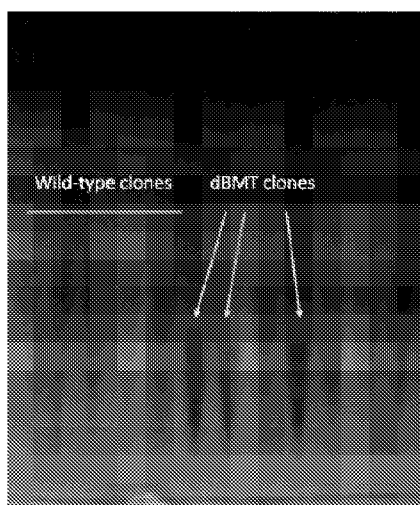


FIG. 1

PROTEIN COMPOSITIONS AND METHODS OF PRODUCTION**CROSS-REFERENCE**

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/225,355, filed July 23, 2021, and U.S. Provisional Patent Application No. 63/356,944, filed June 29, 2022, both of which are entirely incorporated herein by reference.

BACKGROUND

[0002] In industrial protein production, a goal towards cost reduction is to maximize expression of the protein product in the recombinant organism. Methylotrophic yeasts such as *Pichia sp.* are an important production system for proteins. Despite their widespread use, high yield expression, particularly for expression of heterologous animal-derived proteins remains a challenge. This hurdle is particularly apparent in larger scale fermentation settings. While increasing the number of integrated copies can lead to increases in protein expression, there appear to be limitations to the amount of transcript produced with increasing copy number.

[0003] There is a growing demand for animal-free proteins, particularly in food product-based ingredients. For example, an observable trend of preference for health-conscious fast food options has seen egg white demand at all-time highs in recent years. Aside from an increasingly health conscious consumer base, aversion to the inhumane aspects of the industrial hatchery may fuel acceptance and ultimately preference of animal-free egg white alternatives over factory-farmed eggs. Thus, there is a need for novel methods for high-yield industrial production of food proteins, e.g., alternative animal-free egg proteins.

SUMMARY

[0004] In some aspects, provided herein is a recombinant host cell for manufacturing a heterologous protein of interest. In some embodiments, the host cell may be a yeast and may be engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof wherein the underexpression may be compared to the host cell prior to genetic manipulation, wherein the host cell may be engineered to express a heterologous protein of interest and a heterologous mannosidase.

[0005] In some embodiments, the underexpression may be achieved by independently for each mannosyl transferase protein knocking-out the polynucleotide encoding the mannosyl transferase protein or a homologue thereof from the genome of said host cell, disrupting the polynucleotide encoding the mannosyl transferase protein or a homologue thereof in the host cell, disrupting a

promoter which may be operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof, replacing the promoter which may be operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof with another promoter which has lower promoter activity, or disrupting expression control sequences of the mannosyl transferase protein or a homologue thereof, wherein the functional homologue has at least 70% sequence identity to an amino acid sequence of a mannosyl transferase.

[0006] In some embodiments, the host cell may be *Pichia pastoris*.

[0007] In some embodiments, the BMT1 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 12.

[0008] In some embodiments, the BMT2 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 13.

[0009] In some embodiments, the recombinant host cell may be engineered to express at least 10% less BMT1 relative to a host cell which has not been engineered to underexpress BMT1.

[0010] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less BMT1 relative to a host cell which has not been engineered to underexpress BMT1.

[0011] In some embodiments, the recombinant host cell may be engineered to knockout BMT1, wherein the knockout leads to no activity of BMT1 in the recombinant host cell.

[0012] In some embodiments, the recombinant host cell may be engineered to express at least 10% less BMT2 relative to a host cell which has not been engineered to underexpress BMT2.

[0013] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less BMT2 relative to a host cell which has not been engineered to underexpress BMT2.

[0014] In some embodiments, the recombinant host cell may be engineered to knock out BMT2, wherein the knockout leads to no activity of BMT2 in the recombinant host cell.

[0015] In some embodiments, the recombinant host cell produces a reduced size of exopolysaccharides relative to a host cell not engineered to underexpress BMT1 and BMT2.

[0016] In some embodiments, the recombinant host cell may be further engineered to underexpress alpha-1,2-mannosyltransferase MNN2.

[0017] In some embodiments, the MNN2 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 1.

[0018] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNN2 relative to a host cell which has not been engineered to underexpress MNN2.

[0019] In some embodiments, the recombinant host cell may be further engineered to underexpress MNNF1.

[0020] In some embodiments, the MNNF1 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 2.

[0021] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNNF1 relative to a host cell which has not been engineered to underexpress MNNF1.

[0022] In some embodiments, the recombinant host cell may be further engineered to underexpress MNNF2.

[0023] In some embodiments, the MNNF2 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 3.

[0024] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNNF2 relative to a host cell which has not been engineered to underexpress MNNF2.

[0025] In some embodiments, the recombinant host cell may be further engineered to underexpress one or more enzymes in addition to BMT1 and BMT2.

[0026] In some embodiments, the one or more enzymes may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 4-11, 14-15, and 72-85.

[0027] In some embodiments, the recombinant host cell may be engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less one or more enzymes relative to a host cell which has not been engineered to underexpress said one or more enzymes.

[0028] In some embodiments, the recombinant host cell recombinantly expresses a mannosidase from a species different from the recombinant host cell.

[0029] In some embodiments, the mannosidase may be from a genus different from the recombinant host cell.

[0030] In some embodiments, the mannosidase may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 41-56.

[0031] In some embodiments, the mannosidase may be expressed on the surface of the recombinant host cell.

[0032] In some embodiments, the recombinant host cell expresses a surface-displayed fusion protein may comprise a catalytic domain of a mannosidase and an anchoring domain of a glycosylphosphatidylinositol (GPI)-anchored protein, wherein the anchoring domain may comprise at least about 200 amino acids and/or at least about 30% of the residues in the anchoring domain are serines or threonines.

[0033] In some embodiments, the anchoring domain may comprise at least about 225 amino acids, at least about 250 amino acids, at least about 275 amino acids, at least about 300 amino acids, at least about 325 amino acids, at least about 350 amino acids, at least about 375 amino acids, or at least about 400 amino acids.

[0034] In some embodiments, at least about 35% of the residues in the anchoring domain are serines or threonines, at least about 40% of the residues in the anchoring domain are serines or threonines, at least about 45% of the residues in the anchoring domain are serines or threonines, or at least about 50% of the residues in the anchoring domain are serines or threonines.

[0035] In some embodiments, the serines or threonines in the anchoring domain are capable of being O-mannosylated.

[0036] In some embodiments, a fusion protein having an anchoring domain may comprise at least about 325 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain may comprise less than about 300 amino acids.

[0037] In some embodiments, a fusion protein having an anchoring domain may comprise at least about 300 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain may comprise less than about 250 amino acids.

[0038] In some embodiments, the fusion protein may comprise the anchoring domain of the GPI anchored protein.

[0039] In some embodiments, the fusion protein may comprise the GPI anchored protein without its native signal peptide.

[0040] In some embodiments, the GPI anchored protein may be not native to the recombinant host cell.

[0041] In some embodiments, the GPI anchored protein may be naturally expressed by a *S. cerevisiae* cell and the recombinant host cell may be not a *S. cerevisiae* cell.

[0042] In some embodiments, the GPI anchored protein may be selected from Tir4, Dan1, Dan4, Sag1, Fig2, and Sed1.

[0043] In some embodiments, the anchoring domain of the GPI anchored protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 57 to SEQ ID NO: 71.

[0044] In some embodiments, the anchoring domain of the GPI anchored protein may comprise an amino acid sequence of one of SEQ ID NO: 57 to SEQ ID NO: 71.

[0045] In some embodiments, the recombinant host cell may comprise a genomic modification that expresses the fusion protein and/or may comprise an extrachromosomal modification that expresses the fusion protein.

[0046] In some embodiments, the fusion protein may comprise a portion of the mannosidase in addition to its catalytic domain.

[0047] In some embodiments, the fusion protein may comprise substantially the entire amino acid sequence of the mannosidase.

[0048] In some embodiments, the fusion protein, the catalytic domain may be N-terminal to the anchoring domain.

[0049] In some embodiments, the fusion protein may comprise a linker between the catalytic domain and the anchoring domain.

[0050] In some embodiments, the fusion protein may comprise a linker having an amino acid sequence that may be at least 95% identical to any one of SEQ ID NOs: 316-321.

[0051] In some embodiments, upon translation, the fusion protein may comprise a signal peptide and/or a secretory signal.

[0052] In some embodiments, the recombinant host cell may comprise two or more fusion proteins, three or more fusion proteins, or four fusion proteins.

[0053] In some embodiments, the recombinant host cell may comprise a mutation in its AOX1 gene and/or its AOX2 gene.

[0054] In some embodiments, the recombinant host cell may comprise a genomic modification that overexpresses a secreted heterologous protein of interest and/or may comprise an extrachromosomal modification that overexpresses a secreted protein of interest.

[0055] In some embodiments, the secreted protein of interest may be an animal protein.

[0056] In some embodiments, the animal protein may be an egg protein.

[0057] In some embodiments, the egg protein may be selected from the group consisting of ovalbumin, ovomucoid, lysozyme ovoglobulin G2, ovoglobulin G3, α -ovomucin, β -ovomucin,

ovotransferrin, ovoinhibitor, ovoglycoprotein, flavoprotein, ovomacroglobulin, ovostatin, cystatin, avidin, ovalbumin related protein X, and ovalbumin related protein Y.

[0058] In some embodiments, the genomic modification and/or the extrachromosomal modification that overexpresses the secreted recombinant protein may comprise an inducible promoter.

[0059] In some embodiments, the inducible promoter may be an AOX1, DAK2, PEX11, FLD1, FGH1, DAS1, DAS2, CAT1, MDH3, HAC1, BiP, RAD30, RVS161-2, MPP10, THP3, TLR, GBP2, PMP20, SHB17, PEX8, PEX4, or TKL3 promoter.

[0060] In some embodiments, the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein may comprise an AOX1, TDH3, MOX, RPS25A, or RPL2A terminator.

[0061] In some embodiments, the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein encodes a signal peptide and/or a secretory signal.

[0062] In some embodiments, the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein may comprise codons that are optimized for the species of the recombinant host cell.

[0063] In some embodiments, the secreted recombinant protein may be designed to be secreted from the cell and/or may be capable of being secreted from the cell.

[0064] In some embodiments, the additional genomic modification reduces the number of native cell wall proteins expressed by the recombinant host cell, thereby allowing additional space for localization of the surface-displayed fusion protein.

[0065] In some embodiments, the recombinant host cell may comprise a further genomic modification that overexpresses a protein related to the p24 complex.

[0066] In some embodiments, the recombinant host cell may comprise a further genomic modification may comprise that overexpresses more than one protein related to the p24 complex.

[0067] In some embodiments, the protein related to the p24 complex may be selected from Erp1, Erp2, Erp3, Erp5, Emp24, and Erv25.

[0068] In some embodiments, the protein related to the p24 complex may comprise the amino acid sequence of any one of SEQ ID NO: 86 to SEQ ID NO: 91.

[0069] In some aspects, described herein are methods for expressing a heterologous protein of interest. In some embodiments, the method may comprise obtaining a recombinant host cell described herein and culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest.

[0070] In some embodiments, the isolated heterologous protein of interest may be expressed according to the methods described herein.

[0071] In some aspects, provided herein is a method for expressing a heterologous protein of interest. In some embodiments, the method may comprise having of a reduced level of exopolysaccharides, the method may comprise obtaining a recombinant host cell described herein and culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest.

[0072] In some aspects, provided herein is a method for expressing a heterologous protein of interest having of a reduced level of exopolysaccharides. The method may comprise: obtaining a host cell that may be a yeast and may be engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof wherein the underexpression may be compared to the host cell prior to genetic manipulation, wherein the host cell may be engineered to express a heterologous protein of interest and a heterologous mannosidase; and culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest

[0073] In some embodiments, the BMT1 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 12 and the BMT2 protein may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 13.

[0074] In some embodiments, the recombinant host cell may be further engineered to underexpress one or more enzymes may comprise an amino acid sequence of one of SEQ ID NOs: 1-11, 14-15, and 72-85.

[0075] In some embodiments, the recombinant host cell recombinantly expresses a mannosidase from a species different than from the recombinant host cell.

[0076] In some embodiments, the mannosidase may comprise an amino acid sequence that may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 41-56.

[0077] In some embodiments, the mannosidase may be expressed on the surface of the recombinant host cell.

[0078] In some embodiments, the recombinant host cell expresses a surface-displayed fusion protein may comprise a catalytic domain of a mannosidase and an anchoring domain of a glycosylphosphatidylinositol (GPI)-anchored protein, wherein the anchoring domain may

comprise at least about 200 amino acids and/or at least about 30% of the residues in the anchoring domain are serines or threonines.

[0079] In some embodiments, the heterologous protein of interest may be secreted from the recombinant host cell.

[0080] In some embodiments, the secreted heterologous protein of interest may be an animal protein.

[0081] In some embodiments, the animal protein may be an egg protein.

[0082] In some embodiments, the egg protein may be selected from the group consisting of ovalbumin, ovomucoid, lysozyme, ovoglobulin G2, ovoglobulin G3, α -ovomucin, β -ovomucin, ovotransferrin, ovomucoid, ovoglycoprotein, flavoprotein, ovomacroglobulin, ovostatin, cystatin, avidin, ovalbumin related protein X, and ovalbumin related protein Y.

[0083] In some embodiments, the recombinant host cell may comprise a further genomic modification that overexpresses a protein related to the p24 complex.

[0084] In some aspects, provided herein is a method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides. In some embodiments, the method comprises: obtaining a yeast cell engineered to express a heterologous protein of interest and/or a heterologous mannosidase; and modifying the yeast cell to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof.

[0085] In some aspects, provided herein is a method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides. The method may comprise: obtaining a yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous mannosidase; and modifying the yeast cell to express a heterologous protein of interest.

[0086] In some aspects, provided herein is a method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides. In some embodiments, the method comprising: obtaining a yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous protein of interest; and modifying the yeast cell to express a heterologous mannosidase.

[0087] In some aspects, provided herein is a method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides. In some embodiments, the method comprising: obtaining a yeast cell, modifying the yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous protein of interest; modifying the yeast cell to express a heterologous protein of interest; and modifying the yeast cell to express a heterologous mannosidase.

[0088] In some aspects, provided herein are recombinant host cells for manufacturing a heterologous protein of interest. In some embodiments, the host cell may be a yeast cell. The host cell may be engineered to underexpress at least one polynucleotide encoding a mannosyl transferase or a functional homologue thereof compared to the host cell prior to genetic manipulation to achieve underexpression, wherein the host cell is engineered to express a heterologous protein of interest.

[0089] In some embodiments, the underexpression may be achieved by knocking-out the polynucleotide encoding the mannosyl transferase protein or a homologue thereof from the genome of said host cell, disrupting the polynucleotide encoding the mannosyl transferase protein or a homologue thereof in the host cell, disrupting a promoter which may be operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof, replacing the promoter which may be operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof with another promoter which has lower promoter activity, or disrupting expression control sequences of the mannosyl transferase protein or a homologue thereof, wherein the functional homologue has at least 70% sequence identity to an amino acid sequence of a mannosyl transferase.

[0090] In some embodiments, the host cell may be *Pichia pastoris*.

[0091] In some embodiments, the recombinant host cell expresses a mannosidase.

[0092] In some embodiments, the mannosidase may be heterologous to the host cell.

[0093] In some embodiments, the mannosidase may be expressed on the surface of the recombinant host cell.

[0094] In some embodiments, the protein of interest may be a nutritional protein.

[0095] In some embodiments, the mannosyl transferase may be a beta-mannosyl transferase.

[0096] In some embodiments, the beta-mannosyl transferase may be a protein sequence selected from the group consisting of XP_002493882.1, XP_002493883.1, XP_002490760.1, and XP_002493902.1.

[0097] In some embodiments, the mannosyl transferase may be a protein sequence selected from the group consisting of XP_002492593.1, XP_002490149.1, and XP_002493020.1.

[0098] In some embodiments, the host cell may be *Pichia pastoris*.

[0099] In some embodiments, the recombinant host cell expresses a mannosidase.

[0100] In some embodiments, the mannosidase may be heterologous to the host cell.

[0101] In some embodiments, the mannosidase may be expressed on the surface of the recombinant host cell.

[0102] In some embodiments, the protein of interest may be a nutritional protein.

[0103] In some aspects, provided herein are recombinant host cells for manufacturing a heterologous protein of interest. In some embodiments, the host cell may be a yeast cell. The host cell may be engineered to underexpress at least one polynucleotide encoding a protein from the Oligosaccharide Transferase complex or a functional homologue thereof compared to the host cell prior to genetic manipulation to achieve underexpression, wherein the host cell is engineered to express a heterologous protein of interest.

[0104] In some embodiments, the underexpression may be achieved by knocking-out the polynucleotide encoding a protein from the Oligosaccharide Transferase complex or a homologue thereof from the genome of said host cell, disrupting the polynucleotide encoding a protein from the Oligosaccharide Transferase complex or a homologue thereof in the host cell, disrupting a promoter which may be operably linked with said polynucleotide encoding a protein from the Oligosaccharide Transferase complex or a homologue thereof, replacing the promoter which may be operably linked with said polynucleotide encoding a protein from the Oligosaccharide Transferase complex or a homologue thereof with another promoter which has lower promoter activity, or disrupting expression control sequences of a protein from the Oligosaccharide Transferase complex or a homologue thereof, wherein the functional homologue has at least 70% sequence identity to an amino acid sequence of a protein from the Oligosaccharide Transferase complex .

[0105] In some embodiments, the host cell may be *Pichia pastoris*.

[0106] In some embodiments, the recombinant host cell expresses a mannosidase.

[0107] In some embodiments, the mannosidase may be heterologous to the host cell.

[0108] In some embodiments, the mannosidase may be expressed on the surface of the recombinant host cell.

[0109] In some embodiments, the protein of interest may be a nutritional protein.

INCORPORATION BY REFERENCE

[0110] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0111] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0112] **FIG. 1** illustrates the shift in the size of exopolysaccharides using gel electrophoresis after disruption of *BMT1* and *BMT2* genes which suggests that EPS is a form of mannan polysaccharide.

[0113] **FIG. 2** illustrates the growth of *P. pastoris* strains using mannose as a sole carbon source.

[0114] **FIG. 3** illustrates a chromatogram of purified EPS from the parent strain following 2 days of incubation with cells that express surface-displayed mannosidases. The size of the pure EPS byproduct is unchanged following incubation with cells.

[0115] **FIG. 4** illustrates a chromatogram of EPS isolated from Strain 1 cells that express surface-displayed mannosidase enzymes. Strains show no discernable decrease in the concentration of EPS or size of the byproduct molecule.

[0116] **FIG. 5** illustrates a chromatogram of EPS isolated from Strain 2 cell that express the surface-displayed mannosidase enzymes both cause a right shift in the elution profile of the EPS, suggesting a significant change in the size of the polysaccharide molecule.

[0117] **FIG. 6** illustrates size exclusion chromatography of EPS samples. Strain 3 is Strain 1 after the deletion of 5 native *P. pastoris* mannosyltransferases.

[0118] **FIG. 7** illustrates a general schematic for mannosidase surface display.

[0119] **FIG. 8** illustrates size exclusion chromatography of EPS samples. By coupling the deletion of native mannosyltransferases with the expression of a surface-displayed *B. thetaiotaomicron* mannosidase, Strain 4 is able to reduce the size of the EPS byproduct.

[0120] **FIG. 9** illustrates that disruption of native mannosyltransferases is important for B. theta enzymes to recognize mannan as a substrate for cleavage. The strains with deletions and mannosidase elicits the right-shift in the EPS elution profile.

DETAILED DESCRIPTION

[0120] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[0121] High-yielding recombinant protein expression is a cornerstone of various industries such as therapeutic proteins, food industry, cosmetics, etc. Recombinant protein expression though is almost always accompanied by impurities produced by the host cell. Each host cell generates and secretes proteins, carbohydrates, small molecules and polymers that must be separated from the protein of interest (POI) to produce a pure protein composition. The present invention addresses this need. The systems and methods provide high-titer expression of recombinant proteins in large scale production and are particularly useful for expressing pure heterologous animal derived proteins in a microbial host.

[0122] The present invention is concerned with the manipulation of genes related to the production of glycans in host cells. It has been surprisingly found that the manipulated host has an increased capacity to produce a significantly lower amount of exopolysaccharide impurities therefore reducing the amount of impurities produced by the cell while maintaining high-yield of recombinant proteins of interest.

[0123] In a first aspect, the present invention provides a recombinant host cell for manufacturing a protein of interest, wherein the host cell is engineered to underexpress at least one, such as at least 2, or at least 3, polynucleotides encoding a mannosyl transferase, or a functional homologue thereof, wherein the functional homologue has at least 30% sequence identity to an amino acid sequence of these proteins.

[0124] For the purpose of the present invention the term “protein” is also meant to encompass functional homologues of the proteins described.

Knockout (KO) Proteins

[0125] Yeast cells commonly produce highly complex and branched polysaccharides for various purposes such as enforcement for their cell walls. These complex polysaccharides include

mannans with β -1,2-mannosyl linkages. It has not yet been suggested that an alteration in the mannan production pathways may lead to an increased purity of a recombinant protein produced in a yeast or other host cell. Inventors of the current application have discovered for the first time that the underexpression of one or more proteins in the mannosyl transferase pathway and/or the oligosaccharyltransferase (OST) pathway may lead to a reduction in size or amount of the glycans produced by the first cell thereby reducing exopolysaccharide impurities associated with recombinant proteins produced by host cells.

[0126] In some embodiments, a host cell engineered to underexpress one or more KO proteins reduces a concentration of exopolysaccharides produced by the host cell. A decrease in exopolysaccharide concentration can be determined when the exopolysaccharide concentration obtained from an engineered host cell is compared to the concentration obtained from a host cell prior to engineering, i.e., from a non-engineered host cell.

[0127] In some embodiments, a host cell engineered to underexpress one or more KO proteins alters the type of exopolysaccharides produced by the host cell. An alteration in exopolysaccharide concentration can be determined when the exopolysaccharide mass and/or form obtained from an engineered host cell is compared to the mass and/or form obtained from a host cell prior to engineering, i.e., from a non-engineered host cell.

[0128] In some embodiments, one or more proteins from the mannosyl transferase pathway are underexpressed in a host cell. The underexpression of one or more proteins from the mannosyl transferase pathway may lead to a reduced production of mannans in the host cell.

[0129] In one exemplary embodiment, one or more enzymes responsible for forming β -1,2-mannosyl linkages in cell wall mannan may be the KO proteins and may be underexpressed in a host cell. In this example, the mannan structure of the yeast may be altered to produce a reduced amount of the β -1,2-mannosyl linkages. Examples of such proteins include but are not limited to proteins encoded by genes such as BMT2 (SEQ ID NO: 13, XP_002493882.1), BMT1 (SEQ ID NO: 12, XP_002493883.1), BMT3 (SEQ ID NO: 14, XP_002490760.1), and BMT4 (SEQ ID NO: 15, XP_002493902.1), which code for enzymes responsible for forming β -1,2-mannosyl linkages.

[0130] In some embodiments, the host cell may be engineered to underexpress at least one mannosyl transferase enzyme, such as BMT1, BMT2, BMT3 or BMT4. In some embodiments, the host cell may be engineered to underexpress at least two mannosyl transferase enzymes. In some embodiments, the host cell may be engineered to underexpress at least three mannosyl transferase enzymes. In some embodiments, the host cell may be engineered to underexpress at least four mannosyl transferase enzymes.

[0131] In another exemplary embodiment, a host cell may be engineered to express a less complex mannan structure by underexpressing one or more KO proteins. In this example, a protein from the mannosyl transferase pathway, for instance a mannosyl transferase protein may be underexpressed to produce a linear mannan structure with α -1,6-linked mannose units. The α -1,6-linked mannose units may provide for an easier separation from the recombinantly produced POI. Examples of such proteins include but are not limited to proteins encoded by genes such as MNN2 (SEQ ID NO: 1, XP_002492593.1), MNN2 5 homolog 1 (SEQ ID NO: 2, XP_002490149.1), and MNN2 5 homolog 2 (SEQ ID NO: 3, XP_002493020.1).

[0132] In some embodiments, the host cell may be engineered to underexpress two mannosyl transferase enzymes. In one exemplary embodiment, the host cell may be engineered to underexpress BMT1 and BMT2. In one exemplary embodiment, the host cell may be engineered to underexpress one or more enzymes in addition to BMT1 and BMT2. In one example, the host cell may be engineered to underexpress one or more enzymes such as MNN2, MNN2/5 homolog 1 or MNN 2/5 homolog 2 in addition to BMT1 and BMT2.

[0133] In yet another exemplary embodiment, the one or more proteins underexpressed in a host cell may include proteins such as KTR1 (SEQ ID NO: 4, XP_002492424/GQ68_03227T0), KTR1 (alternative start site, SEQ ID NO: 5), KRE2 (SEQ ID NO: 6, XP_002492423/GQ68_03226T0) variant 1, KTR2 (SEQ ID NO: 7, XP_002492102/GQ68_00148T0), KTR3 (SEQ ID NO: 8, XP_002489479/GQ68_02855T0), KTR4 (SEQ ID NO: 9, XP_002490162/GQ68_02152T0), KTR5 (SEQ ID NO: 10, XP_002491999/GQ68_00252T0), MNN4 (SEQ ID NO: 11, XP_002490538/GQ68_01768T0). Exemplary sequences for proteins that can be underexpressed are provided in Table 1. In some cases, the KO protein sequence may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, or at least 99% identical to one or more sequences in Table 1. In some exemplary embodiments, the host cell may be engineered to underexpress one or more enzymes such as KTR1, KRE2, KTR2, KTR3, KTR4, KTR5 and/or MNN4 in addition to BMT1 and BMT2.

[0134] In yet another exemplary embodiment, one or more proteins from the Asparagine Linked Glycolysis (ALG) pathway may be underexpressed in a host cell. In one more exemplary embodiment, one or more proteins from the Oligosaccharyltransferase (OST) may be underexpressed in the host cell. In one or more exemplary embodiments, the proteins in the ALG or OST pathway that may be underexpressed may include a protein with at least 70% identity, at least 75% identity, at least 80% identity, at least 85% identity, at least 90% identity, or at least 95% identity, or at least 99% identity to one or more sequences in Table 7.

[0135] In some embodiments, a host cell engineered to underexpress one or more KO proteins described herein does not negatively impact a yield of the POI produced by the host cell. In some embodiments, a host cell engineered to underexpress one or more KO proteins described herein increases a yield of the POI produced by the host cell. The term “yield” refers to the amount of POI or model protein(s) as described herein, which is, for example, harvested from the engineered host cell, and increased yields can be due to increased amounts of production or secretion of the POI by the host cell. Yield may be presented by mg POI/g biomass (measured as dry cell weight or wet cell weight) of a host cell. The term “titer” when used herein refers similarly to the amount of produced POI or model protein, presented as mg POI/L culture supernatant. An increase in yield can be determined when the yield obtained from an engineered host cell is compared to the yield obtained from a host cell prior to engineering, i.e., from a non-engineered host cell.

[0136] In some embodiments, the host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less KO protein relative to a host cell which has not been engineered to underexpress said KO protein. In some embodiments, the host cell is engineered to knock out the KO protein, wherein the knockout leads to no activity of the KO protein in the host cell.

[0137] In some embodiments, the host cell is engineered to express at most 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% KO protein relative to a host cell which has not been engineered to underexpress said KO protein.

Host Cell

[0138] As used herein, a “host cell” refers to a cell which is capable of protein expression and optionally protein secretion. Such host cell is applied in the methods of the present invention. For that purpose, for the host cell to express a polypeptide, a nucleotide sequence encoding the polypeptide is present or introduced in the cell. Host cells provided by the present invention can be prokaryotes or eukaryotes. As will be appreciated by one of skill in the art, a prokaryotic cell lacks a membrane-bound nucleus, while a eukaryotic cell has a membrane-bound nucleus. Examples of eukaryotic cells include, but are not limited to, vertebrate cells, mammalian cells, human cells, animal cells, invertebrate cells, plant cells, nematodal cells, insect cells, stem cells, fungal cells or yeast cells.

[0139] Examples of yeast cells include but are not limited to the *Saccharomyces* genus (e.g. *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces uvarum*), the *Komagataella* genus (*Komagataella pastoris*, *Komagataella pseudopastoris* or *Komagataella phaffii*), *Kluyveromyces* genus (e.g. *Kluyveromyces lactis*, *Kluyveromyces mandamus*), the *Candida* genus

(e.g. *Candida utidis*, *Candida cacaoi*), the *Geotrichum* genus (e.g. *Geotrichum fermentans*), as well as *Hansenula polymorpha* and *Yarrowia lipolytica*.

[0140] The genus *Pichia* is of particular interest. *Pichia* comprises a number of species, including the species *Pichia pastoris*, *Pichia methanolica*, *Pichia kluyveri*, and *Pichia angusta*. Most preferred is the species *Pichia pastoris*.

[0141] The former species *Pichia pastoris* has been divided and renamed to *Komagataella pastoris* and *Komagataella phaffii*. Therefore, *Pichia pastoris* is synonymous for both *Komagataella pastoris* and *Komagataella phaffii*.

[0142] In some embodiments, the host cell is a *Pichia pastoris*, *Hansenula polymorpha*, *Trichoderma reesei*, *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Pichia methanolica*, *Candida boidinii*, and *Komagataella*, and *Schizosaccharomyces pombe*.

[0143] The terminology used herein is for the purpose of describing particular cases only and is not intended to be limiting.

[0144] As used herein, unless otherwise indicated, the terms “a”, “an” and “the” are intended to include the plural forms as well as the single forms, unless the context clearly indicates otherwise.

[0145] The terms “comprise”, “comprising”, “contain,” “containing,” “including”, “includes”, “having”, “has”, “with”, or variants thereof as used in either the present disclosure and/or in the claims, are intended to be inclusive in a manner similar to the term “comprising.”

[0146] The term “about” or “approximately” means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, e.g., the limitations of the measurement system. For example, “about” can mean 10% greater than or less than the stated value. In another example, “about” can mean within 1 or more than 1 standard deviation, per the practice in the given value. Where particular values are described in the application and claims, unless otherwise stated the term “about” should be assumed to mean an acceptable error range for the particular value.

[0147] The term “substantially” is meant to be a significant extent, for the most part; or essentially. In other words, the term substantially may mean nearly exact to the desired attribute or slightly different from the exact attribute. Substantially may be indistinguishable from the desired attribute. Substantially may be distinguishable from the desired attribute but the difference is unimportant or negligible.

[0148] As used herein, “engineered” host cells are host cells which have been manipulated using genetic engineering, i.e. by human intervention. When a host cell is “engineered to underexpress” a given protein, the host cell is manipulated such that the host cell has no longer the capability to

express the protein described or a functional homologue thereof such as a non-engineered host cell.

[0149] “Prior to engineering” when used in the context of host cells of the present invention means that such host cells are not engineered such that a polynucleotide encoding a knockout (KO) protein or functional homologue thereof is underexpressed. Said term thus also means that host cells do not underexpress a polynucleotide encoding a KO protein or functional homologue thereof or are not engineered to underexpress a polynucleotide encoding a KO protein or functional homologue thereof.

[0150] The term “underexpression” includes any method that prevents or reduces the functional expression of a KO protein or functional homologues thereof. This results in the incapability or reduction to exert its known function. Means of underexpression may include gene silencing (e.g. RNAi genes antisense), knocking-out, altering expression level, altering expression pattern, by mutagenizing the gene sequence, disrupting the sequence, insertions, additions, mutations, modifying expression control sequences, and the like.

[0151] As mentioned herein, a host cell of the present invention is preferably engineered to underexpress a polynucleotide encoding a protein having an amino acid as defined herein. This includes that, if a host cell may have more than one copy of such a polynucleotide, also the other copies of such a polynucleotide are underexpressed. For example, a host cell of the present invention may not only be haploid, but it may be diploid, tetraploid or even more -ploid. Accordingly, in a preferred embodiment all copies of such a polynucleotide are underexpressed, such as two, three, four, five, six or even more copies.

[0152] The terms “underexpress,” “underexpressing,” “underexpressed” and “underexpression” in the present invention refer to an expression of a gene product or a polypeptide at a level less than the expression of the same gene product or polypeptide prior to a genetic alteration of the host cell or in a comparable host which has not been genetically altered. “Less than” includes, e.g., 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80, 90% or more. No expression of the gene product or a polypeptide is also encompassed by the term “underexpression.”

Features of methods of the present disclosure

[0153] In some embodiments, the protein product having a reduced quantity of the exopolysaccharide impurities comprises an at least 50% reduction in exopolysaccharide impurities quantity relative to the composition comprising a recombinant protein of interest and exopolysaccharide impurities. In some cases, the POI product has an at least 75% reduction, at least 80% reduction, at least 90% reduction, or at least 95% reduction in exopolysaccharide

impurities quantity relative to the composition comprising a recombinant POI and exopolysaccharide impurities.

[0154] In various embodiments, less than about 10% of the weight of the POI product comprises the exopolysaccharide impurities. In some cases, less than about 5% of the weight of the POI product comprises the exopolysaccharide impurities.

[0155] In embodiments, the exopolysaccharide impurities (EPS) is generally inseparable from the recombinant POI when using commonly used protein purification methods such as size exclusion chromatography.

[0156] In some embodiments, the EPS component is naturally a component of a recombinant cell's cell wall. In some cases, the EPS present in the composition comprising the recombinant POI was secreted from the recombinant cell rather than being incorporated into the recombinant cell's cell wall.

[0157] In various embodiments, the EPS has an apparent size of about 13kDa to about 27kDa as characterized by a size exclusion chromatography column.

[0158] In embodiments, the EPS comprises mannose. In some cases, the EPS further comprises N-acetylglucosamine and/or glucose.

[0159] In some embodiments, the EPS comprises about 91 mol% mannose, about 5 mol% N-acetylglucosamine, and about 3 mol% glucose as analyzed by gas chromatography in tandem with mass spectrometry. EPS can be quantified using a method using a pb binding column. An analytical HyperREZ XP Pb++ column (8 μ m, 300 x 7.7 mm, Thermofisher Sci.) can be used for the measurement, which is eluted with water on UltiMate 3000 system (Thermofisher Sci.) operated at a flow rate of 0.6 mL/min and monitored with a refractive index detector.

[0160] In various embodiments, the EPS comprises an α (1,6)-linked backbone with α (1,2)-linked branches and/or α (1,3)-linked branches.

[0161] In embodiments, the EPS is a mannan.

[0162] In some embodiments, the recombinant cell is a cell that expresses and/or secretes EPS and is selected from a fungal cell, such as filamentous fungus or a yeast, a bacterial cell, a plant cell, an insect cell, or a mammalian cell.

Methods of Underexpression

[0163] Preferably, underexpression is achieved by knocking-out the polynucleotide encoding the KO protein in the host cell. A gene can be knocked out by deleting the entire or partial coding sequence. Methods of making gene knockouts are known in the art, e.g., see Kuhn and Wurst (Eds.) Gene Knockout Protocols (Methods in Molecular Biology) Humana Press (Mar. 27, 2009).

A gene can also be knocked out by removing part or all of the gene sequence. Alternatively, a gene can be knocked-out or inactivated by the insertion of a nucleotide sequence, such as a resistance gene. Alternatively, a gene can be knocked-out or inactivated by inactivating its promoter.

[0164] In an embodiment, underexpression is achieved by disrupting the polynucleotide encoding the gene in the host cell.

[0165] A “disruption” is a change in a nucleotide or amino acid sequence, which resulted in the addition, deletion, or substitution of one or more nucleotides or amino acid residues, as compared to the original sequence prior to the disruption.

[0166] An “insertion” or “addition” is a change in a nucleic acid or amino acid sequence in which one or more nucleotides or amino acid residues have been added as compared to the original sequence prior to the disruption.

[0167] A “deletion” is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, have been removed (i.e., are absent). A deletion encompasses deletion of the entire sequence, deletion of part of the coding sequence, or deletion of single nucleotides or amino acid residues.

[0168] A “substitution” generally refers to replacement of nucleotides or amino acid residues with other nucleotides or amino acid residues. “Substitution” can be performed by site-directed mutation, generation of random mutations, and gapped-duplex approaches (See e.g., U.S. Pat. No. 4,760,025; Moring et al., *Biotech.* (1984) 2:646; and Kramer et al., *Nucleic Acids Res.*, (1984) 12:9441). Site-directed mutagenesis can be accomplished in vitro by PCR involving the use of oligonucleotide primers containing the desired mutation. Site-directed mutagenesis can also be performed in vitro by cassette mutagenesis involving the cleavage by a restriction enzyme at a site in the plasmid comprising a polynucleotide encoding the parent and subsequent ligation of an oligonucleotide containing the mutation in the polynucleotide. Usually the restriction enzyme that digests the plasmid and the oligonucleotide is the same, permitting sticky ends of the plasmid and the insert to ligate to one another. See, e.g., Scherer and Davis, 1979, *Proc. Natl. Acad. Sci. USA* 76: 4949-4955; and Barton et al, 1990, *Nucleic Acids Res.* 18: 7349-4966. Site-directed mutagenesis can also be accomplished in vivo by methods known in the art. See, e.g., U.S. Patent Application Publication No. 2004/0171 154; Storici et al, 2001, *Nature Biotechnol.* 19: 773-776; Kren et al, 1998, *Nat. Med.* 4: 285-290; and Calissano and Macino, 1996, *Fungal Genet. Newslett.* 43: 15-16. Synthetic gene construction entails in vitro synthesis of a designed polynucleotide molecule to encode a polypeptide of interest. Gene synthesis can be performed utilizing a number of techniques, such as the multiplex microchip-based technology described by

Tian et al. (2004, Nature 432: 1050-1054) and similar technologies wherein oligonucleotides are synthesized and assembled upon photo-programmable microfluidic chips. Single or multiple amino acid substitutions, deletions, and/or insertions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer, 1988, Science 241:53-57; Bowie and Sauer, 1989, Proc. Natl. Acad. Sci. USA 86: 2152-2156; WO 95/17413; or WO 95/22625. Other methods that can be used include error-prone PCR, phage display (e.g., Lowman et al, 1991, Biochemistry 30: 10832-10837; U.S. Pat. No. 5,223,409; WO 92/06204) and region-directed mutagenesis (Derbyshire et al., 1986, Gene 46: 145; Ner et al., 1988, DNA 7:127). Mutagenesis/shuffling methods can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides expressed by host cells (Ness et al., 1999, Nature Biotechnology 17: 893-896). Mutagenized DNA molecules that encode active polypeptides can be recovered from the host cells and rapidly sequenced using standard methods in the art. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide. Semisynthetic gene construction is accomplished by combining aspects of synthetic gene construction, and/or site-directed mutagenesis, and/or random mutagenesis, and/or shuffling. Semisynthetic construction is typified by a process utilizing polynucleotide fragments that are synthesized, in combination with PCR techniques. Defined regions of genes may thus be synthesized de novo, while other regions may be amplified using site-specific mutagenic primers, while yet other regions may be subjected to error-prone PCR or non-error prone PCR amplification. Polynucleotide subsequences may then be shuffled. Alternatively, homologues can be obtained from a natural source such as by screening cDNA libraries of closely or distantly related microorganisms.

[0169] Preferably, disruption results in a frame shift mutation, early stop codon, point mutations of critical residues, translation of a nonsense or otherwise non-functional protein product.

[0170] In another embodiment, underexpression is achieved by disrupting the promoter which is operably linked with said polypeptide encoding the KO protein. A promoter directs the transcription of a downstream gene. The promoter is necessary, together with other expression control sequences such as ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences, to express a given gene. Therefore, it is also possible to disrupt any of the expression control sequence to hinder the expression of the polypeptide encoding the KO protein.

[0171] A nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence on the same nucleic acid molecule. For example, a promoter is

operably linked with a coding sequence of a recombinant gene when it is capable of effecting the expression of that coding sequence.

[0172] In another embodiment, underexpression is achieved by post-transcriptional gene silencing (PTGS). A technique commonly used in the art, PTGS reduces the expression level of a gene via expression of a heterologous RNA sequence, frequently antisense to the gene requiring disruption (Lehtrecek et al., J. Cell Sci (2002). 115:1511-1522; Smith et al., Nature (2000). 407:319-320; Fuhmann et al., J. Cell Sci (2001). 114:3857-3863; Rohr et al., Plant J (2004). 40(4):611-21. Post-transcriptional gene silencing is a biological process in which RNA molecules inhibit gene expression, typically by causing the destruction of specific mRNA molecules using small RNAs including microRNA (miRNA), small interfering RNA (siRNA), or antisense RNA. Gene silencing can occur either through the blocking of transcription (in the case of gene-binding), the degradation of the mRNA transcript (e.g. by small interfering RNA (siRNA) or RNase-H dependent antisense), or through the blocking of either mRNA translation, pre-mRNA splicing sites, or nuclease cleavage sites used for maturation of other functional RNAs, including miRNA (e.g. by Morpholino oligos or other RNase-H independent antisense). These small RNAs can bind to other specific messenger RNA (mRNA) molecules and decrease their activity, for example by preventing an mRNA from producing a protein. Exemplary siRNA molecules have a length from about 10-50 or more nucleotides. The small RNA molecules comprise at least one strand that has a sequence that is “sufficiently complementary” to a target mRNA sequence to direct target-specific RNA interference (RNAi). Small interfering RNAs can originate from inside the cell or can be exogenously introduced into the cell. Once introduced into the cell, exogenous siRNAs are processed by the RNA-induced silencing complex (RISC). The siRNA is complementary to the target mRNA to be silenced, and the RISC uses the siRNA as a template for locating the target mRNA. After the RISC localizes to the target mRNA, the RNA can be cleaved by a ribonuclease. The strand has a sequence sufficient to trigger the destruction of the target mRNA by the RNAi machinery or process is commonly referred to as an antisense strand in the context of a ds-siRNA molecule. The siRNA molecule can be designed such that every residue is complementary to a residue in the target molecule. PTGS is found in many organisms. For yeast cells, the fission yeast, *Schizosaccharomyces pombe*, has an active RNAi pathway involved in heterochromatin formation and centromeric silencing (Raponi et al., Nucl. Acids Res. (2003) 31(15): 4481-4489). Some budding yeasts, including *Saccharomyces cerevisiae*, *Candida albicans* and *Kluyveromyces polysporus* were also found to have such RNAi pathway (Bartel et al., Science Express doi:10.1126/science.1176945, published online 10 Sep. 2009). “Underexpression” can be achieved with any known techniques in the art which lowers gene

expression. For example, the promoter which is operably linked with the polypeptide encoding the KO protein can be replaced with another promoter which has lower promoter activity. Promoter activity may be assessed by its transcriptional efficiency. This may be determined directly by measurement of the amount of mRNA transcription from the promoter, e.g. by Northern Blotting, quantitative PCR or indirectly by measurement of the amount of gene product expressed from the promoter.

[0173] Underexpression may in another embodiment be achieved by intervening in the folding of the expressed KO protein so that the KO protein is not properly folded to become functional. For example, mutation can be introduced to remove a disulfide bond formation of the KO protein or to disruption the formation of an alpha helices and beta sheets.

Protein of Interest

[0174] The term “protein of interest” (POI) as used herein refers to a protein that is produced by means of recombinant technology in a host cell. More specifically, the protein may either be a polypeptide not naturally occurring in the host cell, i.e. a heterologous protein, or else may be native to the host cell, i.e. a homologous protein to the host cell, but is produced, for example, by transformation with a self-replicating vector containing the nucleic acid sequence encoding the POI, or upon integration by recombinant techniques of one or more copies of the nucleic acid sequence encoding the POI into the genome of the host cell, or by recombinant modification of one or more regulatory sequences controlling the expression of the gene encoding the POI, e.g. of the promoter sequence. In general, the proteins of interest referred to herein may be produced by methods of recombinant expression well known to a person skilled in the art.

[0175] There is no limitation with respect to the protein of interest (POI). The POI is usually a eukaryotic or prokaryotic polypeptide, variant or derivative thereof. The POI can be any eukaryotic or prokaryotic protein. The protein can be a naturally secreted protein or an intracellular protein, i.e. a protein which is not naturally secreted. The present invention also includes biologically active fragments of proteins. In another embodiment, a POI may be an amino acid chain or present in a complex, such as a dimer, trimer, hetero-dimer, multimer or oligomer.

[0176] The protein of interest may be a protein used as nutritional, dietary, digestive, supplements, such as in food products, feed products, or cosmetic products. The food products may be, for example, bouillon, desserts, cereal bars, confectionery, sports drinks, dietary products or other nutrition products. Preferably, the protein of interest is a food additive. In some embodiments, the protein of interest is an animal-protein. In some exemplary embodiments, the

protein of interest in an egg-white protein. In some examples, the protein of interest may include one or more proteins such as ovalbumin, ovomucoid, lysozyme ovoglobulin G2, ovoglobulin G3, α -ovomucin, β -ovomucin, ovotransferrin, ovoinhibitor, ovoglycoprotein, flavoprotein, ovomacroglobulin, ovostatin, cystatin, avidin, ovalbumin related protein X, and ovalbumin related protein Y.

[0177] Exemplary POI sequences are provided in Table 5. In some cases, the POI sequence may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, or at least 99% identical to one or more sequences in Table 5.

[0178] In some cases, the protein of interest may be secreted from the host cell.

[0179] In some cases, a POI is produced in a host cell that has been engineered to express or overexpress one or more advantageous protein of interest (APOI). An APOI may be a protein that alters the type or form of glycans produced by the host cell. An APOI may be a protein that reduces glycan production by the host cell. An APOI may be a protein that reduces a type of glycan produced by the host cell. In some embodiments, APOIs may comprise hydrolase enzymes. In one example, APOIs may include mannosyl hydrolases and/or mannosidases. In some examples, the APOIs may comprise one or more helper factor proteins. Examples of such helper factor proteins may include proteins with SEQ ID NOs: 86-91.

[0180] One or more APOIs may be secreted from the host cell using a secretion signal. One or more APOIs may be expressed on the surface of the host cell. APOIs may be expressed on the surface of a host cell using conventional methods of surface display, including but not limited to chimeric linkages of the APOIs with surface display enzymes such as Sed1 (any one of SEQ ID NOs: 64-65), Tir4 (any one of SEQ ID NO: 58-61), Dan1 (any one of SEQ ID NOs: 62-63). Other surface display proteins that may be used are described in Table 4.

[0181] APOIs produced in the host cell may be proteins homologous to the host cell. Alternatively, APOIs produced in the host cell may be heterologous to the host cell. In one example, an APOI comprises a mannosidase such as produced by organisms including the common human gut microbe *Bacteroides thetaiotaomicron*. Exemplary APOIs include proteins with nucleotide sequences in Table 2 (SEQ ID NOs: 16-40) or protein sequences in Table 3 (SEQ ID NOs: 41-56, 86-91). In some cases, the APOI sequence may be at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, or at least 99% identical to one or more sequences in Table 2 or 3.

[0182] In one example, an APOI is a mannosidase which is capable of degrade any of the free altered mannan or exopolysaccharide structures into mannose monosaccharides which the cell can naturally import to use for carbon recovery.

Surface Display of APOIs

[0183] APOIs or the advantageous proteins of interest such as a mannosidase can be displayed on the surface of the host cell. The APOIs displayed on the surface of the cell may be part of a fusion protein.

[0184] In some embodiments, an engineered eukaryotic cell may express a surface-displayed fusion protein comprising a catalytic domain of an APOI and an anchoring domain of a glycosylphosphatidylinositol (GPI)-anchored protein. In some cases, the anchoring domain comprises at least about 200 amino acids and/or at least about 30% of the residues in the anchoring domain are serines or threonines.

[0185] In some embodiments, the anchoring domain comprises at least about 225 amino acids, at least about 250 amino acids, at least about 275 amino acids, at least about 300 amino acids, at least about 325 amino acids, at least about 350 amino acids, at least about 375 amino acids, or at least about 400 amino acids.

[0186] In some embodiments, at least about 35% of the residues in the anchoring domain are serines or threonines, at least about 40% of the residues in the anchoring domain are serines or threonines, at least about 45% of the residues in the anchoring domain are serines or threonines, or at least about 50% of the residues in the anchoring domain are serines or threonines.

[0187] In some embodiments, the serines or threonines in the anchoring domain are capable of being O-mannosylated.

[0188] In some embodiments, a fusion protein having an anchoring domain comprising at least about 325 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain comprising less than about 300 amino acids.

[0189] In some embodiments, a fusion protein having an anchoring domain comprising at least about 300 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain comprising less than about 250 amino acids.

[0190] In some embodiments, the fusion protein comprises the anchoring domain of the GPI anchored protein.

[0191] In some embodiments, the fusion protein comprises the GPI anchored protein without its native signal peptide.

[0192] In some embodiments, the GPI anchored protein is not native to the engineered eukaryotic cell.

[0193] In some embodiments, the GPI anchored protein is naturally expressed by a *S. cerevisiae* cell and the engineered eukaryotic cell is not a *S. cerevisiae* cell.

[0194] In some embodiments, the GPI anchored protein is selected from Tir4, Dan1, Dan4, Sag1, Fig2, or Sed1.

[0195] In some embodiments, the anchoring domain of the GPI anchored protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one or more sequences in Table 4.

[0196] In some embodiments, the anchoring domain of the GPI anchored protein comprises an amino acid sequence of one or more sequences in Table 4.

[0197] In some embodiments, the fusion protein comprises a portion of the APOI in addition to its catalytic domain.

[0198] In some embodiments, the fusion protein comprises substantially the entire amino acid sequence of the APOI.

[0199] In some embodiments, the fusion protein, the catalytic domain is N-terminal to the anchoring domain.

[0200] In some embodiments, the fusion protein comprises a linker between the catalytic domain and the anchoring domain.

[0201] In some embodiments, upon translation, the fusion protein comprises a signal peptide and/or a secretory signal.

[0202] In some embodiments, the engineered eukaryotic cell comprises two or more fusion proteins, three or more fusion proteins, or four fusion proteins.

[0203] In some embodiments, the two or more fusion proteins comprise different enzyme types.

[0204] In some embodiments, the two or more fusion proteins comprise the same enzyme type.

[0205] In some embodiments, the two of the three or more fusion proteins or two of the four or more fusion proteins comprise different enzyme types.

[0206] In some embodiments, the two of the three or more fusion proteins or two of the four or more fusion proteins comprise the same enzyme type. In some embodiments, the three of the three or more fusion proteins or three of the four or more fusion proteins comprise different enzyme types. In some embodiments, the three of the three or more fusion proteins or three of the four or more fusion proteins comprise the same enzyme type. In some embodiments, the each of the two or more, three or more, or four fusion proteins comprise different enzyme types. In

some embodiments, the each of the two or more, three or more, or four fusion proteins comprise the same enzyme type.

Expression of Proteins

[0207] Expression of a recombinant protein such as the POI or the APOI can be provided by an expression vector, a plasmid, a nucleic acid integrated into the host genome or other means. For example, a vector for expression can include: (a) a promoter element, (b) a signal peptide, (c) a heterologous protein sequence, and (d) a terminator element.

[0208] Expression vectors that can be used for expression of a recombinant POI or APOI include those containing an expression cassette with elements (a), (b), (c) and (d). In some embodiments, the signal peptide (c) need not be included in the vector. In general, the expression cassette is designed to mediate the transcription of the transgene when integrated into the genome of a cognate host microorganism.

[0209] To aid in the amplification of the vector prior to transformation into the host microorganism, a replication origin (e) may be contained in the vector (such as PUC_ORIC and PUC (DNA2.0)). To aid in the selection of microorganism stably transformed with the expression vector, the vector may also include a selection marker (f) such as URA3 gene and Zeocin resistance gene (ZeoR). The expression vector may also contain a restriction enzyme site (g) that allows for linearization of the expression vector prior to transformation into the host microorganism to facilitate the expression vectors stable integration into the host genome. In some embodiments the expression vector may contain any subset of the elements (b), (e), (f), and (g), including none of elements (b), (e), (f), and (g). Other expression elements and vector element known to one of skill in the art can be used in combination or substituted for the elements described herein.

[0210] Exemplary promoter elements (a) may include, but are not limited to, a constitutive promoter, inducible promoter, and hybrid promoter. Promoters include, but are not limited to, acu-5, adh1+, alcohol dehydrogenase (ADH1, ADH2, ADH4), AHSB4m, AINV, alcA, α -amylase, alternative oxidase (AOD), alcohol oxidase I (AOX1), alcohol oxidase 2 (AOX2), AXDH, B2, CaMV, cellobiohydrolase I (cbh1), ccg-1, cDNA1, cellular filament polypeptide (cfp), cpc-2, ctr4+, CUP1, dihydroxyacetone synthase (DAS), enolase (ENO, ENO1), formaldehyde dehydrogenase (FLD1), FMD, formate dehydrogenase (FMDH), G1, G6, GAA, GAL1, GAL2, GAL3, GAL4, GAL5, GAL6, GAL7, GAL8, GAL9, GAL10, GCW14, gdhA, gla-1, α -glucoamylase (glaA), glyceraldehyde-3-phosphate dehydrogenase (gpdA, GAP, GAPDH), phosphoglycerate mutase (GPM1), glycerol kinase (GUT1), HSP82, invl+, isocitrate

lyase (ICL1), acetohydroxy acid isomeroreductase (ILV5), KAR2, KEX2, β -galactosidase (lac4), LEU2, melO, MET3, methanol oxidase (MOX), nmt1, NSP, pcbC, PET9, peroxin 8 (PEX8), phosphoglycerate kinase (PGK, PGK1), pho1, PHO5, PHO89, phosphatidylinositol synthase (PIS1), PYK1, pyruvate kinase (pki1), RPS7, sorbitol dehydrogenase (SDH), 3-phosphoserine aminotransferase (SER1), SSA4, SV40, TEF, translation elongation factor 1 alpha (TEF1), THI11, homoserine kinase (THR1), tpi, TPS1, triose phosphate isomerase (TPI1), XRP2, YPT1, and any combination thereof. Illustrative inducible promoters include methanol-induced promoters, e.g., DAS1 and pPEX11.

[0211] A signal peptide (b), also known as a signal sequence, targeting signal, localization signal, localization sequence, signal peptide, transit peptide, leader sequence, or leader peptide, may support secretion of a protein or polynucleotide. Extracellular secretion of a recombinant or heterologously expressed protein from a host cell may facilitate protein purification. A signal peptide may be derived from a precursor (e.g., prepropeptide, preprotein) of a protein. Signal peptides can be derived from a precursor of a protein other than the signal peptides in native a recombinant POI or APOI.

[0212] Any nucleic acid sequence that encodes a recombinant POI or APOI can be used as (c). Preferably such sequence is codon optimized for the species/genus/kingdom of the host cell.

[0213] Exemplary transcriptional terminator elements include, but are not limited to, acu-5, adh1+, alcohol dehydrogenase (ADH1, ADH2, ADH4), AHSB4m, AINV, alcA, α -amylase, alternative oxidase (AOD), alcohol oxidase I (AOX1), alcohol oxidase 2 (AOX2), AXDH, B2, CaMV, cellobiohydrolase I (cbh1), ccg-1, cDNA1, cellular filament polypeptide (cfp), cpc-2, ctr4+, CUP1, dihydroxyacetone synthase (DAS), enolase (ENO, ENO1), formaldehyde dehydrogenase (FLD1), FMD, formate dehydrogenase (FMDH), G1, G6, GAA, GAL1, GAL2, GAL3, GAL4, GAL5, GAL6, GAL7, GAL8, GAL9, GAL10, GCW14, gdhA, gla-1, α -glucoamylase (glaA), glyceraldehyde-3-phosphate dehydrogenase (gpdA, GAP, GAPDH), phosphoglycerate mutase (GPM1), glycerol kinase (GUT1), HSP82, invl+, isocitrate lyase (ICL1), acetohydroxy acid isomeroreductase (ILV5), KAR2, KEX2, β -galactosidase (lac4), LEU2, melO, MET3, methanol oxidase (MOX), nmt1, NSP, pcbC, PET9, peroxin 8 (PEX8), phosphoglycerate kinase (PGK, PGK1), pho1, PHO5, PHO89, phosphatidylinositol synthase (PIS1), PYK1, pyruvate kinase (pki1), RPS7, sorbitol dehydrogenase (SDH), 3-phosphoserine aminotransferase (SER1), SSA4, SV40, TEF, translation elongation factor 1 alpha (TEF1), THI11, homoserine kinase (THR1), tpi, TPS1, triose phosphate isomerase (TPI1), XRP2, YPT1, and any combination thereof.

[0214] Exemplary selectable markers (f) may include but are not limited to: an antibiotic resistance gene (e.g. zeocin, ampicillin, blasticidin, kanamycin, nurseothricin, chloroamphenicol, tetracycline, triclosan, ganciclovir, and any combination thereof), an auxotrophic marker (e.g. *ade1*, *arg4*, *his4*, *ura3*, *met2*, and any combination thereof).

[0215] In one example, a vector for expression in *Pichia* sp. can include an AOX1 promoter operably linked to a signal peptide (alpha mating factor) that is fused in frame with a nucleic acid sequence encoding a recombinant POI or APOI, and a terminator element (AOX1 terminator) immediately downstream of the nucleic acid sequence encoding a recombinant POI or APOI.

[0216] In another example, a vector comprising a DAS1 promoter is operably linked to a signal peptide (alpha mating factor) that is fused in frame with a nucleic acid sequence encoding a recombinant POI or APOI and a terminator element (AOX1 terminator) immediately downstream of a recombinant POI or APOI.

[0217] A recombinant protein described herein may be secreted from the one or more host cells. In some embodiments, a recombinant POI protein is secreted from the host cell. The secreted a recombinant POI may be isolated and purified by methods such as centrifugation, fractionation, filtration, affinity purification and other methods for separating protein from cells, liquid and solid media components and other cellular products and byproducts. In some embodiments, a recombinant POI is produced in a *Pichia* Sp. and secreted from the host cells into the culture media. The secreted a recombinant POI is then separated from other media components for further use.

[0218] In some cases, multiple vectors comprising the gene sequence of a POI and/or APOI may be transfected into one or more host cells. A host cell may comprise more than one copy of the gene encoding the POI and/or APOI. A single host cell may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 copies of the POI and/or APOI. A single host cell may comprise one or more vectors for the expression of the POI and/or APOI. A single host cell may comprise 2, 3, 4, 5, 6, 7, 8, 9 or 10 vectors for the POI and/or APOI expression. Each vector in the host cell may drive the expression of POI using the same promoter. Alternatively, different promoters may be used in different vectors for POI expression.

[0219] A recombinant POI or APOI may be recombinantly expressed in one or more host cells. As used herein, a “host” or “host cell” denotes here any protein production host selected or genetically modified to produce a desired product. Exemplary hosts include fungi, such as filamentous fungi, as well as bacteria, yeast, plant, insect, and mammalian cells. A host cell can be an organism that is approved as generally regarded as safe by the U.S. Food and Drug Administration.

[0220] A host cell may be transformed to include one or more expression cassettes. As examples, a host cell may be transformed to express one expression cassette, two expression cassettes, three expression cassettes or more expression cassettes. In one example, a host cell is transformed express a first expression cassette that encodes a first POI and express a second expression cassette that encodes a second POI.

[0221] The term “sequence identity” as used herein in the context of amino acid sequences is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a selected sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared.

Table 1: Exemplary proteins for underexpression		
SEQ ID NO.	Sequence Info	Amino acid sequence
1	MNN2 (XP_002492593/ GQ68_03403T0)	MFGKRRQVRKLLIWVVLLIVYFFGLQFRAKNSAHQSSIRSFY ADNKEFFDRQYSRYDEYDIIDNMNSHNELLQEQRNGKLAAG LRGVAEEPNSDEVTDDETAIEDEQAAMINFPKRSPQREKSLVEL RKFYKNVLSIIINNKAMPPIENPRDPTPNENALKRKFGKSGIINIA LHDTDPSLPILSEAYLRDSLQLSPSFIASLSKSHSAVVKAFFPSFP ANAYNGTGIVFIGGQKFSWLSLLSIENLRKTGSKVPVELIIPFAH EYEPQLCEEILPKLNATCVLLQETVIGIDLLKSGHLKGYQFKSLA LLASSFEQVLLVDSDNIIVENPDPIFDSEVFQRTGLVLWPDFWR RVTHPDYYKIAGIKLGSEVRHVVDSTYDPSLYTSSSEDPTDIP LHDREGAIPDGSTESGQILISKTKHCQTILLSLYNFFGPDYYYP LFTQGASGEGDKETFLAAANYKLPFYNIKKGVDVIGYWKPD QSAYQGCMLQYDPIVDYQNLQTFLKTHKGSRVNKLQSELD KPGLLSRLIPKFFFRKTFDEHQLQSHFTKDRSKIMFIHSNFPKLD PFGLKLHNYLFVDQDTHKPRIRMYADQTGLSFDLQWIIHE YFCEYPDFNLKYLENANVKPQDLCMFIKEELNFLQNNPIQLT
2	MNN2/5 homolog 1 – MNNF1 (XP_002490149/ GQ68_02166T0)	MLFGLIRHSRRQLLFLGALVTIVLIFTLPNTSPIEANGVKSEEG SITPIIPVLESPANSLEKIVDTASEERIGGATLEEGHENNKEEQAL ENAERAKEKEKTEAIAAEEKLLKAAELLRQQUETTREKEAAKE DDSKKPNQELVEQDQTYLDDIPDDVEDNIISEQDRKKIILPSYTP KTDPAYSKRATALKIFYNDFFIKVADSGPNTAPITKKTRKKGK SKLKGDVSSGDKYEGPVLTEFLRFMEIYSDEFIDAVSESHSKI VNLMPESFPKGMYYQGDGIVIIIGGGVYSWYGLLAIRNLRDGGN TLPVELMLPSDNEYEPQLCEQILPSLNAKCMILSDIVDQDVLKK

		LDFKGYQFKALSLLASSFENVLSLSDSNIPVANVSHLFDHEPFS ETGLVSWPDFWRRTTNPRYYEAAGIKIGEYQVRNCLDGFVPES DFVHIGLKDIPLHDRNGTIPDASTESGQLLVNKNKHAKTLMML FYYNFYGPYYYPLLSQGMAGEGDKETFLAAANFFGLPFYQV KAGPGILGHHDSTGAFTGVAIVQYDPIADYELTKENFVGEKRR GIEAPKAFYGNNNKSPLFHHCNFPKLDPVKLIKEKKLIDNKTH KFNRMYGPNTKLKYDFEERQWKYTKKEYLCEKKYNLLYFTEQ YKNYGQGYSSQERICKFSDRFLKFLSDNPIRIEG
3	MNN2/5 homolog 2 – MNNF2 (XP_00249302 0/GQ68_03863 T0)	MFNSLAPMRLKLLKVFCAVSVLLAATSVVLFHFHGGQIIPIPI ERTVTLSPPANDTWQFQQFFNGYLDALLENNSYPIPERWNH EVTNVRFFNRIGELLSESRLQELIHFSPEFIEDTSDKFDNIVEQIP AKWPYENMYRGDGYVIVGGGRHTFLALLNINALRRAGNKLP VEVVLPTYDDYEEDFCENHFPLLNARCVILEERFGDQVYPRLO LGGYQFKIFAIAASSFKNCFLDSDNIPLRKMDKIFSSELYKNK TMITWPDFWLIRSTSPHYHNITKTPIGDKRVRYFNDFYTNPNE YYYGDEDPRSEIPFHDREGTIPDWTTESGQLVINKEVHFPAILL GLFYNFNGPMGFYPLLSQGGAGEGDKDTFVAASHYYNLPYY QVYKNCESMLYGWVDHANSGRIEHSAIVQYNPIVDYENLQSVK AKAEILKNHEPDSRKKSSKPKSYSKTRLSTHVKGSIYSYRRLF RDSFNKANSDEMFLHCHTPKIEPYRIMEDDLTLGRNKEAKQR WYGGRKNRVRFGYDVELYIWELIDQYICDKNIQYKIFEGKDR DALCGSFMREQLGFLRSTGD
4	KTR1 (XP_00249242 4/GQ68_03227 T0)	MELVRLANLVNVNHPFEQSNIIYRVPLFFLLSTTRPDRTTVQMA GATRINSRVVRFAIFASILVLLGFILSRGSATSYSLPSGLTSDTSQ STGSSPKSESKPSSQGSSGATELKKTYTTDGKEKATFVSLARNS DVWSLASSIRHVEDRFNHFHYDWFVFLNDEEFSDEFKRVTSAL LTSGKAKYGLIPKEHWSFPEWIDKERAACKTRKEMAAKKVIYG DSISYRHMCRFESGFFFRHELMQEYEWYWRVEPDIKIYCDIDY DVFKFMKDNNKMYGFTVSLPEYVATITLWDTTRAFIKENPQ YLPEDNMMDFISDDDGLSYNGCHFWSNFEVGSLSLWRSEAYL KYFDHLDKAGGFFYERWGDAPVHSIAAALFLHRDQIHFFDDV GYFHNPFNNCPVDADLREERRCMCNPKDDFTWKGYSCVPEFF TVNNMKRPKGWEAFSG
5	KTR1 (alternative start site)	SNIYRVPLFFLLSTTRPDRTTVQMAGATRINSRVVRFAIFASILV LLGFILSRGSATSYSLPSGLTSDTSQSTGSSPKSESKPSSQGSSG ATELKKTYTTDGKEKATFVSLARNSDVWSLASSIRHVEDRFN HKFHYDWFVFLNDEEFSDEFKRVTSALTSGKAKYGLIPKEHWS FPEWIDKERAACKTRKEMAAKKVIYGDSISYRHMCRFESGFFFR HELMQEYEWYWRVEPDIKIYCDIDYDVFKFMKDNNKMYGFT VSLPEYVATITLWDTTRAFIKENPQYLPEDNMMDFISDDDGL SYNGCHFWSNFEVGSLSLWRSEAYLKYFDHLDKAGGFFYER WGDAPVHSIAAALFLHRDQIHFFDDVGYFHNPFNNCPVDADL REERRCMCNPKDDFTWKGYSCVPEFFTVNNMKRPKGWEAFSG
6	KRE2 (XP_00249242 3/ GQ68_03226T0) variant 1	MTGCFLNEVPFTDEFKERTSVLISGQAKYGLIPKEHWSYPDYID QERAAESRRQLEDQHVVYGGLESYRHMCRFNSGFFYKHPLML DYRYYYWRVEPEIEILCDVETDLFRYMRENNKTYGFTISIHEFEK TIPTLWETTKEFMKQNPYSIAENLMNFISDDNGKTYNLCHFW SNFEVADMDFWRSADVKEYFKFLDDTGKFFYERWGDAPVHS LAVSLFLPKEKVHFFNEVGKHSVYSMCPIDKDIWKNRKCVC DPNTDFTFRGYSCGRQYYKATGLTRPSNWKDID

7	KTR2 (XP_00249210 2/GQ68_00148 T0)	MKVVWLACFIILAAIWYKDYQSLRGFMDDRVS KTLPIFNAL KLSTNSYIPVDEHLIKPNREP NPKFVKENATLLMLCRNWELEE VLQSMRSLED RFNGRYQYTWTFLNDVPFEKQFIQETTLMASG KTQYALISSTDWNRPSFINETRFEQNLIQSEKDDIYGGSPSYRN MCRFNSGFFYKQKILDQYDYFRVEPGVEYFCDLEEDPFRYM RLHDKKYGFVISL YEYENTIPTLWQTVEKFIENHPEYIHPNNSY EFLTDKEVVGPLGLVALTEQTYNLCHFWSNFEIGDLNFRSEK YEAFFQFLDQAGGFYYERWGDAPVHSIAVGILLDKRQIHFFEN IGYYHLPFSTCPQSYWSYKCNRCICKRNESIDL VPHSCLSKWW KYGGKTF LQ
8	KTR3 (XP_00248947 9/GQ68_02855 T0)	MMRARLSLERNLSFITSVFLASVAVLFISLEMPKVLARDRQIL KLKLGFMGSLQKGSLETSGNIENTESNINSQTTQHIGTIGASN ERANATFYTLCRNEEL YQMLETVQNYEDRFNSKFKYDWVFLN DYPFTDEFKRVISHAISGEAKFGQVPASHWRFPDHIDQQKVYE SMDKMDSDN TTGDYLG LPIPYAKSISYRHMCRYQSGFFYKHG LLQGYKYFWRVEPDVKLYCDIDYDVFKSMEQNGKRYGFVIS MMEFEKTIESLFKEVKNYLQMKGVSRILLED TDNLSD FVYDEL SGDYTLCHFWSNFEIGDLDFRGREYNEFFDYLD SKGGFYER WGD APIHSIAVSLFMQWNDVKWFS DIGYRHPPYLS CPLSEEV R LEKKCSCDPKQDFTMDAYSCTRFYQDIIRD KQKSQGSNP
9	KTR4 (XP_00249016 2/ GQ68_02152T0)	MMISLTKRFTKLAIFGSLSFILTTAGLWLYWD AIQYMMTSGKIP TLDFQFEDFMNRHDDIVDDMMKKYDKIMKAEVKEPNVGNLV YAPESLVDYGRE NATLLMLVRN KELRTALQAIETVESQFNH KF QYPYVFLNDKEFTDKFKSTITEKVSGQVFFETIDKVTWDRPDW IDSAKESERIKVMRKYNVGYADKLSYHNMC RYYSRGFYNHPR LQQFKYYWRFE PGTHYHTSIDYDVFKFMSANDKTYGFVISLY DTERSIETLWPETLKFIEQNPQFVNKNAAWDWLTEKKQNPQK TRIANGYSTCHFW SNFEIGDMDFRSEAYTKWVNHL DATGGF YYERWGDAPVHSIGATLFQDKSKVHWFRDIGYYHAPYYQCP NSPQSDGKCEVGKFSFPNLSDQNCLINWIEMVADNELSMY
10	KTR5 (XP_00249199 9/ GQ68_00252T0)	MSFRLGYIQAIVLGLVLLSVCWTIVIRPD PSSAIDL ASPVTIDLE NSLTNLKSFPISSRRIS SNIDHVFQTGCRNVFKNKKKANAALVV LARNSELEGVQKSMFSMERHFNQWFNYPWIFLNDEEFTESFK DGVMNMTSSGV SFGVISKPDWNFSEEKDRGST EFLRFNEFIQN QGDRGIMYGALPSYHKMCRFYSGYFFKHPLVAKLSWYWRVE PDVEFFCDLTYPFLEMEASGKKYGFAVIKELSNTVPNLFRHT QSFIEKYGISVDEKAW SIFTNRRSFG EKESMKLIDKIRINHLLSN FSGGIGTRLLSSL SRMNLPTS FSSKKPFFY GEEYNLCHFWSNFEI ASTDLFSSPEYESYFQFLEEKKGFYQERWGDAPVHSLAVAMFL NISEIHYFRDIGYRHSNLVHCPKNAPDELQLPYVPASPEYASSA KPDKPPRVSVRDVFRSGRQTEGVNNLNRGSGCRCNC PKKYKE LEDSPSCCIGRWMVL TNDKYKGEKYLDKYSMAEEVKQTL SK GEKLVNKEILKRHHKYPT
11	MNN4 (XP_00249053 8/GQ68_01768 T0)	MKVSKRLIPRRSRLLIMMMLLVVYQLVVLVLGLESVSEGKLA SLLDLGDWDLANSSLSISDFIKLKLKGQKTYHKFDEHVFAAM ARIQSNENGKLADYESTSSKTDVTIQNVELWKRLSEEEYTYEP RITLAVYLSYIHRQTYDRYATSYAPYNLRVPFSWADWIDL TAL NQYLDKTKGCEAVFPRESEATMKLNNITVVDWLEGLCITDKS LQNSVNSTYAE EINSRDILSPNFHVFGYSDAKDNPQQKIFQSKS YINSKLPLPKSLIFLTDGGSYALTVDRTQNK RILKSGLLSHFFSK KKKEHNLPQDQKTF TFDPVYEFNRLKSQVKPRPISSEPSIDSAL

		KENDYKLLKLESSFIFNYGRILSNYEERLESNDFEKSHYESLA YSSLLEARKLPKYFGEVILKNPQDGGIHYDYRFFSGLIDKTQIN HFEDETERKKIIMHRLLRWQYFTYHNNIINWISHGSSLSWYW DGLSFPWDNDIDVQMPIMELNNFCKQFNNSLVVEDVSQGFGR YYVDCTSFLAQRTGRNGNNNIDARFIDVSSGLFIDITGLALTGS TMPKRYSNKLIKQPKKSTDSTGSTPENGLTRNLRQNLNAQVY NCRNGHFYQYSELSPLKLSIVEGALTLPNDFVTILETEYQRRG LEKNTYAKYLYVPELRLWMSYNDIYDILQGTNSHGRPLSAKT MATIFPRLNSDINLKKFLRNDHTFKNIYSTFNVTRVHEEELKHL IVNYDQNKRKSAEYRQFLENLRFMNPIRKDLVTYESRLKALD GYNEVEELEKKQENREKERKEKEKEKEKEKEKEKEKEKEKE EKEKKEKEKEKEKEKEKEKEEYEEDDNEGEQPTQKSQQEAKE
12	BMT1 (XP_00249388 3/GQ68_04782 T0)	MVDLFQWLKFYSMRRLGQVAITLVLLNLFVFLGYKFTPTSTVIG SPSWEPVAVPTVFNESYLDLQFTDINVDSFLSDTNGRISVTCD SLAYKGLVKTSKKKELDCDMAYIRRKIFSSEEYGVLEADLEAQD ITEEQRIKKHWFTFYGSSVYLPEHEVHYLVRRVLFVSKVGRADT PVISLLVAQLYDKDWNELTPHTLEIVNPATGNVTPQTFPQLIHV PIEWSVDDKWKGTEDPRVFLKPSKTGVSEPIVLFNLQSSLCDG KRGMFVTSPFRSDKVNLLDIEDKERPNSEKNWSPFFLDDVEVS KYSTGYVHFVYSFNPLKVIKCSLDTGACRMIYESPEEGRFGSEL RGATPMVKLPVHLSLPKGKEVWVAFPRTRLRDCGCSRTTYRP VLTLFVKEGNKFYTELISSSIDFHIDVLSYDAKGESCSGSISVLIP NGIDSWDVSKKQGGKSDILTLTLSEADRNTVVVHVKGLLDYL LVLNGEGPIHDSHSFKNVLSTNHFKSDTTLLNSVKA AECAIFSS RDYCKKYGETRGEPARYAKQMENERKEKEKEKEKEAKEKLEA EKAEMEEAVRKAQEAIAQKEREKEEAEQEKKAQQEAKEKEAE EKA AKEKEAKENEAKKKIIVEKLAKEQEEAEKLEAKKKLYQL QEEERS
13	BMT2 (XP_00249388 2/GQ68_04781 T0)	MRTRLNFLLLCIASVLSVIWIGVLLTWNDNNLGGISLNGGKDS AYDDL LSLGSFNDMEVDSYVTNIYDNAPVLGCTDLSYHGLLK VTPKHDLACDLEFIRAQILDIDVYSAIKDLEDKALTVKQKVEK HWFTFYGSSVFLPEHDVHYLVRRVIFSAEGKANSPVTSIIVAQI YDKNWNELNGHFLDILNPNTGKVQHNTFPQVLPIATNFVKGK KFRGAEDPRVVLRLKGRFGPDPLVMFNSLTQDNKRRRIFTISPFD QFKTVMYDIKDYEMPRYEKNWVPFFLKDNQEA VHFVYSFNPL RVLKCSLDDGSCDIVFEIPKVDSMSSELRGATPMINLPQAIPMA KDKEIWVSFPRTRIANCGCSRTTYRPMMLFVREGSNFFVELL STSLDFGLEVLPPYSGNGLPCSADHSVLIPNSIDNWEVVDSDNGD DILTLSFSEADKSTSVIHIRGLYNYLSELDGYQGPEAEDEHNFQ RILSDLHFDNKT TVNNFIKVQSCALDAAGYCKEYGLTRGEAE RRRRVAERKKKKEKEEEEEKKKKKEKEEEEEKKRIEEKKKIEEK ERKEKEKEEAERKKLQEMKKKLEEITEKLEKGQRNKEIDPKEK QREEERKERVRKIAEKQRKEAEKKEAEKKANDKKDLKIRQ
14	BMT3 (XP_00249076 0/ GQ68_01534T0)	MRIRSNVLLLSTAGALALVWFVAVVFSWDDKSIFGIPTPGHAVA SAYDSSVTLGTFNDMEVDSYVTNIYDNAPVLGCYDLSYHGLL KVSPKHEILCDMKFIRARVLETEAYAALKDLEHKKLTEEEKIE KHWFTFYGSSVFLPDHDVHYLVRRVVFSGEGKANRPITSILVA QIYDKNWNELNGHFLNVLNPNTGKLQHHAFPQVLPIAVNWDR NSKYRGQEDPRVVLRRGRFGPDPLVMFNLTQTQNNKLRLFTIS PFDQYKTVMYRTNAFKMQTTEKNWVPFFLKDDQESVHFVYS

[illegible]

76	CCW22 homolog GQ68_03003 (chr 3)	MSLLLFLVLGAFLSSVKAADIGAFRLRVYTPGRFTNGALNFN NWGYQYLDASSNGQLFAGYATVTSVTTFLLAPDDEGFVWGSS LGGYPGFLGIGAGATAFHLTGIPGDALSWYIEDNILKTSSPTYV CSRNDGDVVVGIEANTRWLAMHDTSQLPPNYYCFQADYEIVA LWYIPDTTSTWTGTETSTTTDDDDGSVIELVPTPLPDTTSTWTGT FTTFTTDDDDGSVIELVPTPLPDSTSTWTGTYYTFTTDEDGSTIAV VPSSTIDSTSTWTGTYYTFTTDEDGSTIAVPSSTIDSTSTWTGT YTTFTTDEDGSTIAVYHHLLSTPHPPGLVLTTPRSLPMRMEVLLL WYHHLLSTLHPPGLVLTTPRSLPMRMEVLLLWYHRLSTPHPG LVLTPRSLPMRMEVLLL YHHLLSTPHPPGLVLTTPRSLPMRMEV LLLWY
77	FLO5 homolog GQ68_04296 (chr 4)	MKLQLQSFVFFLLSAVNVLADDSYGCSIATSPRSTGFVANLYE FPNMAISNAELKTYVRVRYKEGRLYDTISNIISPYFYYQGQGA NSAYGTLYGRPNVYLYNFSMELKGYFRPPITGQYTIDFNGANV DDAAMVFFGKAGAFDCCNSDYILPEQSAEYSLYSVYPHTATD QILSATIYLEAGKYYPLRVITYTNIGNIGSLDLRVVLP SGASITSL GAFVYQFPNNLSPGTCTPDVEYFTTTTQAWTGTYYETTYTVPPS GTQPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTGTEPGTVIIET TPESYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQ PWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYET YTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEP GTVVIETPEITDCEAVCCGAVPTSDPLRRRDVDCDCETFCPCGDT NCETYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTT QPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYET YTVPPTGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEP GTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIETPESY VTTTQPWTGTYYETTYTVPPTGTEPGTVIIETPESYVTTTQPWTG TYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPP TGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTGTEPGTVII ETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTT QPWTGTYYETTYTVPPSGTEPGTVVIETPEITDCEAVCCGAVPTS DPLRRRDVDCDCETFCPCGDTNCETYVTTTQPWTGTYYETTYTV PPSGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTGTEPGTV IIETPESYVTTTQPWTGTYYETTYTVPPSGTQPGTVIIETPESYVTT TQPWTGTYYETTYTVPPTGTEPGTVIIETPESYVTTTQPWTGTYE TTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGT QPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTGTEPGTVIIETPE SYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWT GTYYETTYTVPPTGTEPGTVIIETPESYVTTTQPWTGTYYETTYTV PPTGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTV IIETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTVVIETPEITDCE AVCCGAVPTSDPLRRRDVDCDCETFCPCGDTNCETYVTTTQPW TGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYETTYT VPPTGTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTQPGT VIIETPESYVTTTQPWTGTYYETTYTVPPTGTEPGTVIIETPESYV TTQPWTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGT ETTYTVPPSGTQPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTG TEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTVIIET ESYVTTTQPWTGTYYETTYTVPPTGTEPGTVIIETPESYVTTTQP WTGTYYETTYTVPPSGTEPGTVIIETPESYVTTTQPWTGTYYETTY TVPPSGTQPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEPG

		TVIVETPDVPGSYVTTTQPWTGTYYETTHTVPPTGTTEPGTVVVE TPDVPGSYVTTTQPWTGTYYETTHTVPPTGTTEPGTVVVE TPDVPGSYVTTTQPWTGTYYETTYTVPPSGTEPGTVIVETPDVPGSYVTT TQPWTGTYYETTHTVPPTGTTEPGTVVVE TPDVPGSYVTTTQPWT GVYKTTYTVPPSGTIPGTVIIETPFQYFNTSSISTKTDKRTITSVV PCSQCSESKTQYITPTGPGDVTVIIISQPPSKITLSSPEDKTKTDFI TSTGSIGGGSPPSHPNDKPGIITPTQPIGGGNPSDIPSAISSVSSG GNSRASVPSFSTSSAISVQVSSLYDENSGSTFEVSLLFVSVVSGFF LTLMV
78	FLO5 homolog GQ68_03011 (PAS_chr3_114 5)	MKFPVPLLFLQLFFIIATQGDSENGNGDESDTAYGCDITSNAFD GFDATIIYEYNANDLKLIRDPVFMSTGYLGRNVLNKISGVTVP G FNIWNPRSRATATVYGVQNVNYYNMVLELKG YFKA AVSGDYK LTL SNIDDSSMLFFGKNTAFQCCDTGSIPVDQAPTDYSLFTIKPS NQVNSEVISSTQYLEAGKYYPVRIVFVNALERALFNFKLTIPSG TVLDDFQDYIYQFGALDENSCYETTVSKITEWTTYYTTPWTGTF ETTRITPTGTGEGTVVIETPESYVTTTQPWTGTYYETTYTVPPTGT EPGTVIIETPEIIDCEAVCCGPFLTAFSFRKREECQCENICCPGDT NCETYVTTTQPWTGTYYETTYTVPPTGTTEPGTVIIETPESYVTTT QPWTGTYYETTYTVPPTGTTEPGTVIIETPESYVTTTQPWTGTYYET TYTVPPSGTEPGTVVIETPEIVDCEAYCCASVAIKKRELCQCEN FCCSWDQSCQTYVTTTQPWTGTYYETTYTVPPTGTTEPGTVIIETP ESYVTTTQPWTGTYYETTYTVPPTGTTEPGTVIIETPESYVTTTQP WTGTYYETTYTVPPTGTTEPGTVIIETPEIIDCEAVCCGPFLTAFS FRKREECQCENICCPGDTNCETYVTTTQPWTGTYYETTYTVPPTGT TEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPTGTTEPGTVIIETP ESYVTTTQPWTGTYYETTYTVPPTGTTEPGTVIIETPEIINCEAVCC GPFLTAFSFRKREECQCENICCPGDTNCETYVTTTQPWTGTYYE TTYTVPPTGTTEPGTVIIETPESYVTTTQPWTGTYYETTYTVPSTGT EPGTVIIETPESYVTTTQPWTGTYYETTFVTPPTGTTEPGTVVIETP ESYVTTTQPWTGTYYETTYSVPPSGTEPGTVVIETPEASTARTKF TTVTSSWTGVFTTTKTL PASGTEPATIVIQPTGYFNTSSLVSTR TKTNVDTVTRVIPCPICTAPKTITVPEEPNESVSVIISQPQSSST DTTLSKPDSVRVISQPETASQMDTSLSKTDSAVISTETAGNNIIP LAGSHSYNTIVTTVTDSPQVAQSTTATSSSNVHLTISTQTTTPSL VYSSSLSTVHQVSPSNGGFRSSITVHPLLSVIGAIFGALFM
79	FLO5 homolog GQ68_03079 (chr 3)	MTKFTILLVLLKFYSILAEVDGSA NGQPLAHPIVVEVHEATK WITH TSPWTGTPEAIRTVTGETPYEQKIARYDEFNPRLANREIID CVAFC CGDATSSPSITEPESTATELPESYVTINRPWSLSWIPDVP PGSPYWSTSTIPPSGTEPGTVIIFYLYDDARKRREINFGSTQPY HGRPKLLGSIEKRELCQCD AVCCGLDLSCEVYVTTTQPWTGT YETTYTITPTGSEPGTVIIETPELYVTTTQPWTGTYYETTYTITPTG SEPGTVIIETPESYVTTTQPWTGTYYETTYTITPTGSEPGTVIIETP ESYVTTTQPWTGTYYETTYTITPTGSEPGTVIIETPESYVTTTQPW TGTYYETTYTVPPSGTEPGAVIIETPELYVTTTQPWTGTYYETTYTI TPTGSEPGTVIIETPESYVTTTQPWTGTYYETTYTVPPSGTEPGTV IIETPELYVTTTQPWTGTYYETTYTITPTGSEPGTVIVEIPVSYVNS TQISTSTYD TTDVTLSSGVEPGTIAIETPIVYLN TSVSAFSRPWT KIDTVTQFSSCAVCSKPETITVTPENPIDTVTHIISQPQSTSQSNTPT TSFKANSTSAFSRFEDEDSIPVFGSYSYEITVNIDVNTEDDTTNL NADTTIIIGSLSAIRTVAGSSSNYHASNISPTINSQKTASSVVVHS DSSATVYOFSPSNGAPWLSVOISTLLSVVGTL LAAVLL

80	FLO5 homolog GQ68_04277 (chr 4)	MNFYLLILPIYASIVLGQVGDFQLLLNAKEPIRNSPSLLSSNYG NLTLPAMANGALESHFDYGNAYVGDDQITVVYHLPDEHGQIN AYRQDTDEYIGYLGLVTDDYGEYTYLSVIMPGVQYDQTTSVN WYIENEELKSTSINVQPLLCYYKNPPQYSWYWASIDEPGNIA SSNFVCEPCKVYVDFVPTPSADITSMWTGSETSWTTDADGTVI ELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQIPTPSADITSM WTGSETSWTTDADGTVIELVPTPSADITSMWTGSETSWTTDAD GTVIELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQIPTPSADI TSMWTGSETSWTTDADGTVIELVPTPSADATSVWTGDHTTWTT TDDDDGNVIEQIPTPSADITSMWTGSETSWTTDADGTVIELVPTP SADITSMWTGSETSWTTDADGTVIELVPTPSADATSVWTGDHT TWTTDDDDGNVIEQIPTPSADITSMWTGSETSWTTDADGTVIEL VPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADITSMWT GSETSWTTDADGTVIELVPTPSADATSVWTGDHTTWTTDDDDG NVIEQIPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADITS MWTGSETSWTTDADGTVIELVPTPSADITSMWTGSETSWTTD ADGTVIELVPTPSADITSMWTGSETSWTTDADGTVIELVPTPSA DITSMWTGSETSWTTDADGTVIELVPTPSADITSMWTGSETSW TTDADGTVIELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQIP TPSADITSMWTGSETSWTTDADGTVIELVPTPSADITSMWTGS ETSWTTDADGTVIELVPTPSADATSVWTGDHTTWTTDDDDGNV IEQIPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADATSV WTGDHTTWTTDDDDGNVIEQIPTPSADITSMWTGSETSWTTDA DGTVIELVPTPSADITSMWTGSETSWTTDADGTVIELVPTPSAD ITSMWTGSETSWTTDADGTVIELVPTPSADATSVWTGDHTTW TTDDDDGNVIEQIPTPSADITSMWTGSETSWTTDADGTVIELVPT PSADITSMWTGSETSWTTDADGTVIELVPTPSADATSVWTGDH TTWTTDDDDGNVIEQIPTPSADITSMWTGSETSWTTDADGTVIE LVPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADITSMW TGSETSWTTDADGTVIELVPTPSADATSVWTGDHTTWTTDDDD GNVIEQIPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADIT SMWTGSETSWTTDADGTVIELVPTPSADITSMWTGSETSWTTD ADGTVIELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQIPTPS ADITSMWTGSETSWTTDADGTVIELVPTPSADITSMWTGSETS WTTDADGTVIELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQ IPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADITSMWTG SETSWTTDADGTVIELVPTPSADITSMWTGSETSWTTDADGTV IELVPTPSADATSVWTGDHTTWTTDDDDGNVIEQIPTPSADITSM WTGSETSWTTDADGTVIELVPTPSADITSMWTGSETSWTTDAD GTVIELVPTPSADITSMWTGSETSWTTDADGTVIELVPTPSADA TSVWTGDHTTWTTDDDDGNVIEQIPTPSADITSMWTGSETSWTT DADGTVIELVPTPSADITSMWTGSETSWTTDADGTVIELVPTPS ADITSMWTGSETSWTTDADGTVIELVPTPSADTTSVWTGSYTT WTTDEDGTVIEQVPTPSADTTSVWTGSYTTWTTDEDGTVIEQVPTPSADTP SADTTSVWTGSYTTWTTTEVGDDGSSTVVELVPTESSTSTNVM QTPVPSSGVSDGVSFNGFNVEVFHYPADNYELANEISFLSYG YENLGLVTTVTGVSDINFDTDSNWPYYIDRDALGNTGSYVNA TIEYEGFFRAPVDGEYVFSFSSTDYNSILFVGSPAAADQALQKR EVQFLKPETSPDYVLLFNNTDRLGKTVSTTQYLLADQYYPLRV VIAAISQHALLDFQIKLPNGASLTQYQGYVYNFALEGSESTTVI
----	---------------------------------------	---

		GDKTSTWTGSYTTWTTDSDGSTIVVPPATITADKTSTWTGSY TTWTTDSDGSTVICPSITSDHNDKPSESTLTDSSISTTVVTVTS CDIEKCTKTTALTGVRETTLTGGTTTVVTTYCPLPTDIVTVKT TSIDGSEVLQTIYTAKPNHVVPDVQTSTVTITREVCDAFTCTHA TIVTGEILKTTTLADTHYTTVVPVYVPLETYQPAVELSTLETVL KSSDLASGPVVTAGSVQPSYQSGGVAESSLTVSEFEAHSTSDT VSQPSTISLQTGEANALKWSSFFGAALVPLVNVFFV
81	FLO5 homolog GQ68_01371 (chr 1)	MQNTNDKLIIRTFYSISTIHGLLSINIFSDTRVYKFAIYSTDAVSL EPRTKNNMSLVTVLACFIIFAAHAFGQDTFYMLKVRTLTPNGY PLADSLSNPMQYWDLYYVPGGPRRLESSFVNWQPTTAAPINQ FYCRLGTDGHMTGYNRVTGSGVIGKLSFGTNAATALAFGSYDG DPSYPPQAFSISSSVSGTMTYLVNHYVNARSITWYSTTTATGET NVYINVASTGYTGDRTTYQAEWVEPFVNPVDTTTSIWTS QTSYTTEVGENGSTVIELIPTPPADATSTWTGTYTTRTTDADG SVIEQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTT VWTGTYTTWTTDADGSVIEQIPTPSADTTATWTGTETSYTTDV GEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVV ELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADT TATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTET SYTTDVGEDGSSTVVELVPTPSADTTATWTGTETSYTTDVGED GSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVP TPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTAT WTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYT TDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSS TVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPS ADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTG TETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDV GEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIEL VPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPTPSADT TATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTET SYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGED GSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVP TPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTA TWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSY TTDVGEDGSSTVVELVPTPTPSADTTATWTGTETSYTTDVGED GSSTVVELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVP TPSADTTATWTGTETSYTTDVGEDGSSTVVELVPTPSADTTAT WTGTETSYTTDVGEDGSSTVVELVPTPTADTTATWTGTETSYT TDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSS TVVELVPTPSADTTATWTGTETSYTTDVGEDGSSTVVELVPTP SADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWT GTETSYTTDVGEDGSSTVVELVPTPSADTTATWTGTETSYTTD VGEDGSSTVVELVPTPTADTTATWTGTETSYTTDVGEDGSSTV IELVPTPSADTTATWTGTETSYTTDVGEDGSSTVVELVPTPSAD TTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTE TSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGE DGSSTVIELVPTPTPSADTTATWTGTETSYTTDVGEDGSSTVIE LVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPTPSAD TTATWTGTETSYTTDVGEDGSSTVVELVPTPSADTTATWTGTE TSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGE DGSSTVVELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIEL

		VPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTA TWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSY TTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGS STVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPS ADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSADTTATWTG TETSYTTDVGEDGSSTVIELVPTPSADTTATWTGTETSYTTDVG EDGSSTVIELVPTPSADTTATWTGTETSYTTDVGEDGSSTVIEL VPTPSADTTATWTGTETSYTTDVGEDGSSTVIELVPTPSAD TTATWTGTETSYTTDVGEDGSSTVIELVPSDTETATNIVETPVP SSGVSDGVSFDFGNVEVFHYPADNYELANEIGFLSYGYENLG LVTNATGVSDINFDTDSNWPYYIDRDALGNTGSYVNATIEYEG FFRAPVDGEYVFSFNTDYNLSILFVGSPAAAGQALQKRRVQFL KPETSPDHVLLFNNTDRLGQTISTTQYLLADQYPLRVVIAAIS QHALLDFQIKLPNGALLTQYQGYVYNFALEGSESTTVIGDKTS TWTGSYTTWTTSDSGSTVVVPSATITADKTSTWTGSYTTWT TDSGSGTIVICPSITSDHNDKPSESTLTDGSISTTVVTVTSCDIEK CTKTTALTGVTEETLTGTTTTVVTTCPLPTDIVTVKTTSSISGS EVLQTIYTAKPSHVVPNVHTLTVTITREVCDRAFTCTQATIVTGE ILKTTTLADTHSTTVVPVYVPLESYQSAVELSTLETVLKSSDFA SGSAVTAGSAQPSYQSGGVAESSLTGSELEAHSTSDTVSQPSTI SPQTGEANALRWSSFFGAALVPLVNVFFV
82	FLO5 homolog GQ68_04678 (PAS_chr4_036 3)	MTKLILLSVLLQLFSVLAIEVPKKTEWSSHTTYWTSTLEALRT VTPTGTERAVIGEAPYEYKLGINDQFDPGLNAKREIIDCEAVCC GAVPTSDPLKRRDVCECENVCCPGDDCETYVTTTQPWTGTYE TTYTVPPSGTEPGTVVIETPEITDCEAVCCGAVPTSDPLRRRDV CECENVCCPGDDCETYVTTTQPWTGTYEETTYTVPPSGTEPGTV VIETPEITDCEAVCCGAVPTSDPLRRRDVCECENVCCPGDDCE TYVTTTQPWTGTYEETTYTVPPSGTEPGTVVIETPTVYVTTTQP WTGTYEETTYTVPPSGTEPGTVVIETPEITDCEAVCCGAVPTSDP LRRRDVCECENVCCPGDDCETYVTTTQPWTGTYEETTYTVPPS GTEPGTVVIETPTVYVTTTQPWTGTYEETTYTVPPSGTEPGTV ETPTVYVTTTQPWTGTYEETTYTVPPSGTEPGTVVIETPTVYV TTPWTGTYEETTYTIPPTGTEPGTVVIETPEITDCEAVCCGAVPT DPLRRRDVCECENVCCPGDDCETYVTTTQPWTGTYEETTYTV PTGTEPGTVVIETPTVYVTTTQPWTGTYEETTYTVPPSGTEPG VIETPTVYVTTTQPWTGTYEETTYTVPPSGTEPGTVVIETPTV YVTTTTPWTGTYEETHTVPASGTEPGTVIETPIKYLNTSISAST STWTKINTVTQFISCPVCTIPKTITVTPKISNETVTIISQPHGT SSRTTTVKTDGASVSSHYSYKTALTTDVKPEEKTSTKLGTVTTV SGSHSAIDTVTGSLSDYHASSIPHTVKSEEKASSTVHTISSSTV YQVSPSNGASWLSVRLNTALSIIGTLFAAVFI
83	FLO5 homolog GQ68_04282 (chr 4)	MSKTKNGGSEFVHIAVVFHIEASTPSDYINMIQIVLFPHQAQIT KRMNLVTLVCNLLCVSLTLGQGVYRLKFPALVVTGRESVGT TVVNYDFLVGNTGQYGDLEFFYDGEPPYYCWNSTDSQPLSCS SSSSLLISTQNVTISHPDEDGTVYAYAERDGGLLGRFTVGSVSA DWPQWAVIVYSTSSSAHPSSWYVDDNKLKLTSGLGPNNSTTL QACYFTQSSGRDRYAISLEGSPAYTGQVSCQATEFDLEFIPPSA DTTSIWDGSYTTWTTSDNGIVVEQIPTPSADTTSIWTGSETSWT TDSGTVIELVPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPS ADTTSIWTGSETSWTTSDGTVIELVPTPSADATSIWTGDHTT WTTDSEGNVIEQIPTPSADATSIWTGDHTTWTTDREGNVIEQIP

		TPSADTTSIWTGSETSWTTSDSGTVIELVPTPSADATSIWTGDH TTWTTDSEGNVIEQIPTPSADATSIWTGSETSWTTSDSGTVIEL VPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPSADATSIWTG DHTTWTTDSEGNVIEQIPTPSADTTSIWTGSETSWTTSDSGTVI ELVPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPSADATSIW TGDHTTWTTDSEGNVIEQIPTPSADTTSIWTGSETSWTTSDSGT VIELVPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPSADATSI WTGDHTTWTTDSEGNVIEQIPTPSADTTSIWTGSETSWTTSDS GTVIELVPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPSADA TSIWTGSETSWTTSDSGTVIELVPTPSADATSIWTGDHTTWTT DSEGNVIEQIPTPSADATSIWTGDHTTWTTDSEGNVIEQIPTPSA DTTSIWTGDHTTWTTTEVGGDGSSIVVELVPSETGTATNVVQTP VPSSGISDGVSA LDGFNVEVFHYPADNYELANEISFLSYGYENL GLVTTATGVSDINFD TDSNWPSYIDRNALGNTGSYVNATIKYE GFFRAPVDGDYEF SFSNIDYNSILFVGSAADQALRKREAQFL KPETSPNHILFFNNSRDVGQTISTTQYLSADSYPLRVVIAAVS QHALLDFQIKLPNGVSLTQFQGYVYNFALEGAESTTVIGDKTS TWTGTYTTWTTDSEGSTIVLCPSIISDHNGKPADTTLDGSISTT VVTVTSCDIKKCTKTAL TGVTQKTLTVKGT TTVTAYCPLPT DVATVKTISVGGSEVLQTVYTA KP SHIVPDVQTLTVTITREVC DALT CIPATIVTGEILKTTTLADTHSTTVIPVYVPLETHQPALDLIT LETVLKSSDFANGPAITSVSVESLSHQSGVVVSEFDS DSTSGAV SQPSSAVSLQTGKASALKWSPFLGAAVISLFNVFFV
84	FLO5 homolog GQ68_03013 (PAS_chr3_001 5)	MNLFTILAWGFLYVPLVLGEGYYSLNFDARVPIALGILGSSYQ KYTIMADRSLGGSNIDLDTFSGIHELLTNRVHIVVSLPADG RVSVDYDMYSGTSLGYLSFVCSLTTCVHAVSSSSGATTWTL GNQLIPTSPSTVYACYRSLVGLLAQYTLNDRTSITAQCEQTNLY VELAIPAFPETTAVWTGTYTTWTTDESGSVIEQMPTPSADTTT WTGTYTTWTTDADGSVIEQIPTPPADTTSVWTGTYTTRTTDAD GSVIEQIPTPSADTTSIWTGTYTTWTTDADGSVIEQIPTPSADTT SVWTGTYTTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTT DADGSVIEQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPS TDTTLAPSADTTSIWTGTYTTWTTDADGSVIEQIPTPSADTTSI WTGTYTTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTTDA DGSVIEQIPTPSTD TTLAPSADTTSIWTGTYTTWTTDADGSVIE QIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTTSVWT GTYTTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTTDADGS VIEQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTTSV WTGTYTTWTTDADGSVIEQIPTPSADTTSIWTGTYTTWTTDAD GSVIEQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTT LAPSADTTSIWTGTYTTWTTDADGSVIEQIPTPSADTTSIWTGT YTTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTTDADGSVI EQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTTSVW TGTYTTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTTDADG SVIEQIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTT VWTGTYTTWTTDADGSVIEQIPTPSTD TTLAPSADTTSIWTGTY TTWTTDADGSVIEQIPTPSADTTSVWTGTYTTWTTDADGSVIE QIPTPSADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTTSVWT GTYTTWTTDADGSVIEQIPTPSTD TTLAPSADTTSIWTGTYTT TTDADGSVIEQIPTPSADTTSIWTGTYTTWTTDADGSVIEQIPT SADTTSVWTGTYTTWTTDADGSVIEQIPTPSADTTSVWTGTYT

		TWTTDADGGSVIEQIPTPSADTTSVWTGTYTTWTTDADGGSVIEQI PTPSADTTSVWTGTYTTWTTDADGGSVIEQSPTPSAYTTSVWTG TYTTWTTDADGGSVIEQIPTPSADTTSVWTGTYTTWTTDADGGSV IEQIPTPSADTTSVWTGTYTTWTTDADGGSVIEQIPTPSADTTSV WTGTYTTWTTDADGGSVIEQIPTPSADTTSVWTGTYTTWTTDA DGGSVIEQIPTPSADTTSVWTGTYTTWTTDADGGSVIEQIPTPSAD TTSVWTGTYTTWTTDADGGSVIEQSPTPSAYTTSVWTGTYTTW TTDADGGSVIEQIPTPSADTTLAPSADTTSIWTGTYTTWTTDADG SVIEQIPTPSADTTSIWTGTYTTWTTDADGGSVIEQIPTPSADTT VWTGTYTTWTTDADGGSVIEQIPTPSADTTSVWTGTYTTWTTD ADGGSVIEQIPTPSADTTSVWTGTYTTWTTDADGGSVIEQIPTPS DTTLAPSADTTSIWTGTYTTWTTDADGGSVIEQIPTPSADTTSVW TGTYTTWTTDADGGSVIEQIPTPSDITLAPSADTTSIWTGTYTT WTTDADGGSVIEQIPTPSADTTSVWTGTYTTWTTDADGGSVIEQIP TPSADTTSVWTGTYTTWTTDADGGSVIEQIPTPSADTTSVWTGT YTTWTTDADGGSVIEQIPTPSADTTLAPSADTTSIWTGTYTTWTT DADGGSVIEQIPTPSADTTSVWTGTYTTWTTDAAGTVIEVIPS GTSISSDVIPTPLPTSGVDITIPYDAFNVA VYHYPADNYELAN NLGFLTSGYEGLGQVTTATSVGNINFDTS SGPYIESNALGNTG SYVNATIEYVGFFQAPANGNYELSFSNIDYNAILFLGSPATDSS LAKREVQFLKPETSSEYVLFFDHGKDAGQTVSTTQYLSAGLY YPLRIVLAAVSERAQLDFQITLPDGRVLDQYQGYVYNFAHEGI ESATSSAHETSWSRFTNSTIYSHSSTIGIITSSTDAPHSVINPTAIE TTSTDTSISTVAVTTSICDTKDCVKTTVITPNSPLPTQTVSLTTT IDRSEVVQTAHSAVPSQFAPDAHPSAVTITREQCDAYSCSQATI VSGKVLQTTTVSDSTTVVPLDTPQLSVEASTLETRLKSTQSSRA PTVTVQTSQSSRHSEDTVTESSVHVSEFDAQSTSATSASALQAPS SISLQTGGANTLRLSAFLGTALLPMLNVLF
85	SED1 homolog (GQ68_01572)	MQFSIVATLALAGSALAAYSNVTYTYETTITDVVTELTTCPEP TTFVHKNKTITVTAPTTLTITDCPCTISKTTKITTDVPPTTHSTPH TTTTHVPSTSTPAPTHSVSTISHGGAAGVAGLAGVAAAAA YFL

Table 2: Exemplary advantageous proteins (Nucleotides)

SEQ ID NO.	Sequence Info	Nucleotide sequence
16	BT2623 Bacteroides thetaitaom icron mannan utilization genes	Native nucleotide: ATGAAAAAAGTAATAAAGAAATATTTCTTTTTAGCATTAGCTAT TATAATGTATTCGTGTAATGAAGATGAAAAGTATGATATATTAG AAAGATACACTCCTGAACTATAACATCTGACGAAATAGCTCCT GTGCTTAATTTACAGGCACAATATATGGATAGTAATAGCGAAAT AGTACTTGTAACATGGATGAATCCGGAAGATGATTTTTTGTCTA AGGTGGAAATCTCTTGCTGTTCTGCGAATGATAATCTTTGGGT GAACCTGTGTTGTTGGACGCTGTTTCTACCAAAGTAGGTTCTTAT CAGACGTCACCTTTCTGTGGAAGAGAGGGGATATGTAAAGATTGT AGCTATTAATGAAAAAGGAGTACGCTCGGAAGCCCGTACAGCA GAGATCCTTTCTTCCCAACAGGATTTTGTATATAGAGCAGATTG TTTGATGTCTTCTGTGATTGAATTATTTTTTGGTGGGAGATATAA TGCATGGAATGAGAATTACCCCAATGCTACAGGTCCCTATTGGG

		ATGGCATTGCAGCCGTTTGGGGACAAGGTGCAGCTTATTCCGGA TTTGTACAATGTATAAGGTCACAAAGGAACTAATAATGAGAA ACTAAGAGCAAAATATGCAGAAAAGGAAGAACTTTTCTAAAC TCAATAGACATTTTTTTTGAATAATGGTAGTGGACGGAAATCTTT TGCTTATGGTACTTATATTGGGCCGAATGATGAGCGTTATTACG ATGATAATGTCTGGATTGGCATCGAAATGGCCAATTTATATGAA CTTACAGGGAATGAAGTTTATTTGCAGCATGCAAATACTGTTTG GAACTTTATTTTGGAAAGGGATAGATGACGTGACTGGCGGTGGAG TATATTGGAAAGAAGGTGCGGTATCAAAGCATACATGTTCCACT GCCCCGGCAGCTGTAATGGCTCTAAAATTATACCAATTGAGCAA GAATGAATCATATTTGGAAATAGCAAAGAGTTTGTATTCATACT GTAAAGATGTATTACAAGATCCGAATGACTATTTATTTTATGAC AATGTTTCGCTTAAGTGACCCTTCCGATAAGAATTCGGAGCTTAA AGTATCTAAGGATAAATTCACGTATAATTCGGGACAACCAATGT TAGCTGCTGCTATGTTGTATCGGATTACAAAAGAAGAACAATTT CTGAAAGATGCCCCAAAATATAGCACAGTCGATTTATAAAAAAT GGTTTAAAACTATCATTCGTCTATACTTGATAGAGATATAATG ATATTAAGCGATCCAAACACTTGGTTTAATGCCGTTATGTTTCAG GGGATTCGTAGAGCTATATAAAATAGATAAGAACGATGTTTATG TCAAAGCGGTGAAAAATACCATGGAACATGCTTGGCAAAGCAA CTGTAGAAATCGGTTGACTAATCTAATGAGCGACGATTATGCAG GTGATAAGAAAGAAGGTAAGTGGAATATAAAGACACAAGGTGC TTTTGTTGAAATCTTCTCACTTATTGGGGAATTGGAACAACCTGG ATGTTTTTCAGGAGTAG
17	BT2623 (codon optimized) Bacteroides thetaitaom icron mannan utilization genes	ATGAAGAAAGTAATTAAGAAATATTTTTTCTAGCCTTGGCAAT CATTATGTACTCATGTAACGAAGACGAGAAATATGACATTCTTG AACGTTATACCCCTGAAACTATAACCTCTGACGAGATCGCACCT GTACTAAACCTTCAAGCCCAGTACATGGATTCAAACAGTGAAAT AGTTCTTGTGACTTGGATGAACCCAGAGGATGATTTTCTGAGTA AAGTTGAGATTTCTTGCTGCAGTGCTAACGATAACTTACTGGGTG AGCCCGTCCTTCTTGATGCCGTCTCAACCAAGGTCGGCTCCTAC CAGACGTCCCTTTCTGTGCAAGAACGTGGATATGTTAAGATCGT AGCTATAAATGAAAAGGGAGTTAGGTCTGAGGCTAGGACGGCT GAGATTTTGTGATCTCAACAAGACTTCGTCTATCGTGCAGACTG CCTTATGTCTAGTGTGATTGAACTGTTCTTTGGAGGAAGGTACA ATGCATGGAACGAAAATTACCCCAATGCAACCGGCCCTTACTGG GATGGAATCGCCGCTGTGTGGGGTCAGGGTGCAGCCTATTCTGG TTTCGTAACATATGTACAAAGTTACCAAAGAAACAAATAACGAA AACTAAGGGCTAAGTATGCAGAAAAGGAGGAAACATTCTCTGA ACTCTATAGACATCTTTTAAATAATGGCTCTGGCAGAAAGTCA TTTGCCTACGGCACGTACATCGGTCCTAACGACGAGCGTTATTA CGATGATAATGTGTGGATAGGTATAGAAATGGCAAACCTTATATG AGCTGACAGGAAACGAGGTGTACCTACAACATGCCAATACCGT GTGGAATTTTCATATTAGAAGGCATTGATGATGTAACGGGAGGTG GCGTATACTGGAAGGAGGGTGCAGTTTCCAAACACACGTGCTCA ACCGCCCCCGCAGCTGTAATGGCTTTGAACTTTACCAGTTGTC CAAGAATGAATCCTACTTAGAGATCGCCAAATCCTTGTATTCTT ACTGCAAAGATGTCTTGCAAGATCCAAACGATTATCTTTTTTAC GACAACGTGAGGCTAAGTGACCCTTCAGATAAGAACAGTGAAC TAAAAGTATCAAAAGACAAGTTCCTTACAACAGTGGTCAGCCC ATGCTTGCAGCAGCCATGCTGTATCGTATAACCAAAGAAGAGCA

		GTTTCTGAAAGACGCCCAAACATTGCCCAATCAATATACAAGA AATGGTTCAAAAATTACCATTCAATCTTAGATAGGGATATA ATGATTTTGTCTGATCCAAACACCTGGTTTAACGCAGTCATGTTT AGGGGTTTTGTCTGAGCTGTATAAAATCGACAAAAATGATGTTTA TGTTAAGGCAGTTAAGAACACAATGGAGCATGCTTGGCAATCA AACTGCCGTAACAGACTTACCAATCTTATGTCTGACGACTATGC CGGAGACAAGAAGGAGGGTAAGTGGAACATTAAGACCCAAGG AGCTTTTGTGAAATTTTTCTTTGATTGGCGAGTTAGAACAGTT AGGCTGTTTCCAGGAATAG
18	BT2629	ATGAAAACACATTTTTCATTTAACACTTGCTATTTCTTGGAGGT GCGGTGTTGTACAGCCTGCAATCTTCTGCCGTTAAGAATCCTGT AGACTATGTCAGCACACTGATAGGCACTCAATCCAAGTTTGAAC TGTCTACCGGAAACACGTATCCGGCTACGGCATTGCCGTGGGGA ATGAATTTCTGGACACCGCAGACCGGTAAAATGGGAGACGGTT GGGCGTACACGTATGATGCCGACAAAATCCGGGGATTCAAACA AACACATCAGCCCAGTCCCTGGATGAACGACTACGGGCAGTTTCG CCATCATGCCTATCACAGGCGGACTGGTATTCGATCAAGACCGA CGTGCCAGTTGGTTCTCTCACAAAGCGGAAGTTGCCAAACCTTA TTATTATAAGGTATACCTCGCCGACCATGATGTAACAACCGAGC TTGCTCCTACGGAGCGTGCCGTCATGTTCCGTTTCACGTATCCGG AGACAAAGAATGCCTACGTGATTGTAGACGCTTTCGACAAAGGT TCTTATGTGAAAGTGATTCCGGAAGAAAACAAGATTATCGGCTA TTCAACCAAGAATAGCGGCGGTGTGCCGGAACCTTCAAAAAC TATTCGTGATTCAATTCGACAAACCGTTCACATTCGTTTCCACA GTTTTCGAAAACAACATTCTTCCGAATGAAACAGAAGCAAAAAG GAAACCACACAGGGGGCGTGATCGGATTCGCCACGAAAAAGGG AGAAATCGTACACGCACGTGTTGCTTCCTCCTTTATCAGCCCCG AACAGGCGGAGTTGAATCTCAAAGAGCTTGGCAAAAACAGTTT CGACCAACTGGTAGCGAACGGAAGAGAAATCTGGAACCGTGAA ATGAGTAAAATAGAGATAGAAGACGATAATATCGATAATTTAC GCACCTTCTATTCTTGTTTATACCGTTCCATGCTTTTTCCACGCA GTTTCTACGAGATAGATGCTAAGGGACAAGTCATGCATTACAGC CCCTACAACGGCGAAGTGCGTCCCGGTTATATGTTTACCGACAC CGGATTCTGGGACACGTTCGCTGCCTGTTCCCTTTCCTCAACCT GATGTATCCGTCAATGAATCAAAAGATGCAGGAGGGACTAGTG AATACTTACAAGGAAAGTGGTTTCCCTGCCGGAATGGGCCAGTCC GGGACATCGGGATTGTATGGTAGGCAACAACCTCGGCTTCCGTAG TAGCCGACGCTTACATCAAAGGATTGCGAGGATATGATATCGAA ACTCTTTGGGAAGCATTGAAACATGGAGCAAATGCACATCTTCG CGGGACTGCTTCAGGTCGTCTCGGTTACGAATCTTACAACCAAC TGGGATATGTTGCCAACAATATCGGCATAGGACAAAACGTTGCA CGTACATTGGAGTATGCTTACAACGACTGGGCAATTTATACACT AGGTAAGAACTTGGTAAACCGGAGAACGAAATCGACATTTAT AAGAAACACGCGCTGAACTACAAAAATGTCTATCACCCGGAAC GCAAACCTGATGGTTGGCAAAGATAACAAAGGCGTATTCAATCC GAATTTTCGATGCAGTGGACTGGAGCGGTGAATTTTGCGAAGGG AATAGCTGGCACTGGAGCTTCTGCGTATTCCACGACCCGCAAGG ACTTATCAACCTGATGGGAGGCAAGAAAGAATTCAACGCGATG ATGGATTCTGTTTTTGTCTATCTCGGGTAACTGGGAATGGAAAG CCGCGGCATGATTCACGAAATGCGTGAAATGCAAGTAATGAAC ATGGGGCAATATGCGCATGGCAACCAGCCTATTCAACACATGGT

		ATATCTCTACAAC TATTCAAGCGAAC CCTGGAAAGCTCAATACT GGATACGTGAGATTATGAACAACTATATACCGCCGGTCCCGAC GGTTATTGCGGTGACGAAGACAACGGACAGACTTCCGCCTGGTA TGTATTCTCCGCACTCGGTTTCTATCCGGTTTGGCCGGGAACAGA TGAATATATCATAGGAACCCCGCTCTTTAAATCAGCGAAGTTAC ATTTGGAGAACGGAAAGACCATCACGATCAAGGCAGATAACAA CCAGCTTGACAACCGCTACATCAAGGAAATGAAAGTAAACGGG AAATCACAAACCCGTAATTTCTTACACATGACCAGCTGATTAA AGGTGCTAATATTCAATTTCAAATGAGCCCCGTGCCCAATAAAC AACGGGGAACACAGAAAAAGATGTACCTTACTCTCTTTCGTTT GAATAA
19	BT2629 (codon optimized)	ATGAAAACACATTTTTCATTTAAACACTTGGCTATTTCCTTGGAGGT GCCGTGTTGTACAGCCTGCAATCTTCTGCCGTAAATAATCCCGT CGACTATGTGTCTACCCTTATAGGCACGCAATCCAAGTTTGAGT TGTCCACAGGCAACACCTACCCTGCTACCGCTCTTCCATGGGGC ATGAACTTTTGGACTCCACAGACAGGAAAAATGGGTGATGGAT GGGCATATACGTACGATGCTGACAAGATCCGTGGCTTTAAACAA ACTCACCACCATCTCCATGGATGAACGACTACGGTCAGTTTGC AATAATGCCAATTACTGGAGGACTTGTATTTGACCAAGATAGAC GTGCTAGTTGGTTTTCCACAAAGGCAGAAGTCGCTAAACCATAC TATTACAAGGTCTACCTTGCTGACCATGACGTGACAACCGAATT GGCCCCACCGAGAGGGCCGTGATGTTTAGGTTTACGTACCCCG AGACGAAAAACGCCTACGTTATTGTAGATGCCTTTGATAAGGGA AGTTATGTCAAAGTAATACCTGAGGAAAACAAGATTATAGGTTA TTCTACAAAAAATTCAGGCGGCGTCCCAGAAAATTTTAAGAACT ACTTCGTTATTCAGTTTGACAAACCATTTACGTTTCGTATCAACTG TATTTGAAAACAATATTTTGCCAAACGAGACAGAGGCCAAGGG TAACCACACAGGCGCTGTGATCGGCTTCGCAACGAAGAAGGGC GAAATAGTACATGCTAGAGTCGCTCTTCTTTCATATCTCCTGAA CAAGCCGAGTTAACTTAAAGGAATTGGGAAAAAATTCTTTTGA TCAACTGGTAGCCAACGGTAGGGAGATTTGGAATCGTGAGATG AGTAAGATCGAGATCGAGGATGATAACATTGATAATTTAAGGA CGTTCTATTCTTGTCTGTATAGATCCATGTTGTTTCTAGGTCCTT TTACGAGATTGACGCTAAGGGCCAGGTGATGCACTATTCACCCT ACAATGGCGAAGTACGTCCTGGATACATGTTACGGGATACGGG ATTTTGGGACACGTTTAGGTGTCTGTTCCCTTTTTTGAATCTGAT GTATCCCTCCATGAACCAGAAAATGCAGGAGGGCCTTGTAAC ACTTACAAGGAGTCCGGATTTTTTACCAGAGTGGGCAAGTCCAGG CCATCGTGATTGTATGGTTGGCAACAATTCAGCATCAGTTGTGG CTGATGCCTATATCAAAGGTTTGAGAGGATACGATATCGAGACG CTGTGGGAGGCCCTTAAACACGGTGCCAACGCTCATCTAAGGGG TACCGCATCTGGCAGATTAGGTTACGAGTCCTACAACCAACTAG GCTACGTGGCTAATAATATCGGTATTGGCCAGAACGTTGCAAGA ACCCTTGAATACGCTTACAACGACTGGGCAATCTACACTTTGGG TAAAAAACTTGGAACCCGAAAATGAAATAGACATTTATAAG AAACACGCTCTTAACTACAAAAACGTGTATCACCTGAAAGGA AGCTAATGGTCGGTAAGGACAACAAGGGCGTCTTTAACCCTAAT TTCGATGCTGTGGACTGGTCTGGAGAGTTCTGCGAAGGCAATTC CTGGCATTGGTCCTTCTGTGTTTTTCACGACCCTCAAGGATTAAT TAATTTGATGGGTGGTAAGAAGGAGTTCAATGCTATGATGGATT CCGTATTCTGTGATCTCTGGTAACTGGGCATGGAGTCTCGTGGT

		ATGATCCACGAAATGAGAGAGATGCAGGTAATGAACATGGGAC AATACGCACATGGCAATCAGCCTATACAGCATATGGTATATCTT TATAACTACAGTTCAGAGCCTTGGAAGGCACAATATTGGATTAG GGAGATCATGAACAAGCTTTATACCGCCGGCCCTGATGGATATT GTGGCGATGAAGATAACGGACAGACCAGTGCATGGTATGTGTTT TCCGCACTTGGTTTTTACCCTGTGTGCCCTGGTACGGATGAGTAC ATTATCGGCACGCCATTATTCAAATCTGCTAAGTTGCATCTTGA AAACGGAAAGACGATAACGATAAAAGCCGACAACAACCAACTG GATAACAGATATATTAAAGAAATGAAGGTCAACGGTAAGTCAC AAACGAGAACTTTTTAACCCTATGACCAACTAATTAAGGGAGCC AACATACAATTCCAGATGAGTCCAGTCCCCAATAAGCAACGTGG AACAAACAGAGAAGGACGTGCCTTATTCTTTGTCCTTCGAGTAG
20	BT2630	ATGAAACTGAAAAACCTTTTACTAATTGCCCTTGTTGCGATCGT CTTTTGCGGTGTCAAAGTAACTATCAGCCTACTTCTATCACCGT TGCCTCCTACAATTTGAGAAACGCCAACGGTGGCGATTCAATCA ACGGAAACGGTTGGGGACAACGTTACCCGGTCATTGCCCAAAT AGTGCAATATCACGATTTTCGATATTTTCGGCACGCAGGAGTGCT TTATTCATCAACTGAAAGATATGAAAGAAGCATTACCCGGTTAT GATTATATCGGTGTAGGTCGCGACGACGGCAAAGAGAAAGGTG AACATTCTGCTATTTTCTATCGCACAGACAAGTTTGACGTGATA GAGAAAGGTGATTTTTGGTTGTCGGAAACTCCCGACGTGCCGAG CAAAGGATGGGATGCCGTGTTGCCGCGTATTTGCAGTTGGGGAC ACTTCAAATGCAAAGATACCGGCTTCGAATTCCTTTTCTTCAACC TGCACATGGACCATATCGGCAAGAAGGCACGTGTGGAAAGTGC ATTCTCGTACAGGACAAGATGAAAGAAGTGGCAAAGGCAAA GAGCTTCCGGCCATCCTGACGGGAGACTTCAATGTCGACCAGAC CCACCAGTCTTATGATGCTTTTGTGAGCAAAGGGGTGTTGTGCG ACTCTTACGAGAAGGCCGGCTTCCGCTATGCTATCAACGGCACG TTCAACGACTTCGACCCGAACAGCTTTACGGAAAGCCGTATCGA CCATATATTCGTTTCTCCGTCTTTCCAAGTGAAAAGATATGGTGT GCTGACTGATACTTACCGCAGCATCGTAGGCAAGGGAGAAAAG AAGCAGGCGAACGATTGCCCGGAAGAAATCGACATCAAGACTT ATCAGGCGCGCACTCCTTCAGACCATTTCGCCGTAAAGGTGGAA CTGGAGTTCGACCAGCGTCAGCAGAAATAA
21	BT2631	ATGAGAAATATATGTTTTGTAGCCTGTATGTTATTTTGCCCTACT TCCGCACTGGGAAAGACACCGGGAAATACCCGTTATCTTTCTAT TGCCGACTCGATTCTATCTAATGTATTGAATCTCTATCAGACGA ATGACGGACTACTAACAGAAACGTATCCTGTCAATCCCGACCAA AAAATTACTTATCTGGCGGGCGGAACGCAGCAGAACGGAACGC TGAAGGCTTCTTTTCTATGGCCGTATTCCGGGATGATGTCGGGT GTGTGGCTTTATACAAAGCGACCGGAAACAAGAAGTACAAAAA GATTCTCGAGAAAAGAATTCTACCGGGAATGGAGCAGTATTGG GATAACAGTCGCTTGCCGGCCTGTTATCAGTCATACCCACCAA GTACGGGCAGCACGGACGTTATTATGACGATAACATCTGGATTG CACTGGATTACTGCGATTATTACCAACTGACTCACAAGCCTGCA TCTTTGGAAAAAGCCGTTGCATTGTATCAATATATCTACAGTGG ATGGAGCGATGAGATAGGCGGTGGCATCTTTTGGTGTGAACAGC AGAAGGAAGCGAAGCATACTTGTTCCAATGCACCGTCTACTGTG CTCGGTGTCAAGTTGTACCGGCTGACGAAGGATGCCAAATACCT CGAAAAAGCAAAAGAGACGTATGCCTGGACGAAAAAGCATCTG TGCGACCCTACCGACCATCTTTACTGGGATAACATCAACCTGAA

		AGGGAAAGTTTCCAAAGAGAAGTACGCCTACAACAGTGGACAG ATGATTCAGGCGGGTGTATTGCTCTATGAGGAAACGGGAGATG AACAGTATTTGCGCGATGCACAGCAGACAGCCGCAGGAACTGA TGCTTTTTTCCGCACAAAAGCCGACAAGAAAGACCCGACTGTCA AAGTGCATAAAGACATGGCCTGGTTTAACGTGATCTTATTCAGA GGACTGAAAGCTCTGTATAAGATTGACAAGAATCCGGCGTATGT CAATGCGATGGTGGAAAATGCGCTTCACGCCTGGGAAAACACTAC CGGGATGAAAACGGATTATTAGGCAGGGATTGGTCGGGACATA ACAAGGAGCAGTATAAATGGCTGCTCGACAATGCCTGTCTTATT GAATTCTTTGCAGAGATTTAA
22	BT2631 (codon optimized)	ATGAGAAACATCTGCTTTGTCCCTGTATGCTGTCTGTCTGACC AGTGTCTGTGGGCAAGACTCCTGGAAACACGAGGTACCTATCTAT TGCCGACTCTATCCTTTCCAACGTGTTGAACCTTTACCAAATAA CGATGGTCTTCTGACCGAACTTATCCTGTTAACCCCGACCAGA AGATAACCTATTTGGCTGGCGGCACACAACAGAATGGCACCCCTG AAGGCATCTTTTTTGTGGCCTTATTCTGGCATGATGTCCGGATGC GTTGCATTGTATAAAGCCACTGGCAACAAAAAGTATAAAAAGA TACTTGAGAAAAGGATTTTACCAGGAATGGAGCAGTACTGGGA CAATAGTCGTTTACCAGCATGTTATCAATCATAACCCTACTAAAT ACGGCCAGCACGGAAGATACTATGACGATAATATCTGGATCGC CTTAGATTACTGCGACTATTACCAGTTAACCACAAACCCGCCT CTCTGGAGAAAGCCGTAGCTCTATATCAGTACATCTATTCTGGT TGGTCAGATGAGATTGGCGGAGGCATATTTTGGTGTGAGCAACA AAAAGAGGCCAAGCACACGTGCTCCAATGCCCTTCCACTGTAT TAGGTGTAAAACTGTATAGGCTTACAAAAGACGCCAAATATCTG GAAAAAGCTAAAGAGACGTATGCTTGGACCAAGAAACATCTTT GCGACCCTACAGATCATTTGTACTGGGATAATATAAACTTGAAA GGAAAGGTTTCTAAAGAAAAATACGCCTATAATAGTGGTCAAA TGATTCAGGCCGGCGTTCTGTTGTATGAGGAAACAGGCGATGAG CAATATCTTCGTGATGCTCAACAAACAGCCGCTGGCACAGACGC ATTTTTCAGAACGAAGGCAGACAAGAAAGACCCAAGTGTCAAG GTACATAAGGACATGGCCTGGTTTAACGTAATTTTATTTAGAGG CCTGAAGGCATTATATAAAATAGACAAGAACCCCGCCTATGTAA ATGCTATGGTAGAGAATGCCCTGCATGCCTGGGAAAATTACAGA GACGAGAATGGACTTCTAGGAAGAGATTGGAGTGGTCACAACA AAGAACAATACAAATGGCTATTAGATAACGCCTGTCTAATTGAG TTCTTCGCAGAGATTTAG
23	BT2632	ATGAATATAACTAAAGCCTTTTGTGTTGTCCATAGCACTCTTGGGC GCTAGCAATATGCAGGCTATAACGAACAGTGATTTTGTTCATCCA ACAAGATAATACCAAAATCAACAACTATCAGACGAACCGTCCG GAAACATCGAAACGTCTGTTTGTCTCACAAGCTGTGGAACAACA GATTGCGCATATCAAGCAACTGCTGACGAATGCCCGCTTAGCAT GGATGTTTCGAAAACGTGTTCCCGAACACACTGGATACTACTGTT CATTTTGACGGTAAAGACGATACGTTTGTGTTATACAGGTGACAT CCACGCCATGTGGTTGCGCGATTCCGGGTGCACAAGTATGGCCTT ACGTGCAACTCGCCAACAAAGACGCAGAACTGAAAAAAATGCT CGCTGGCGTTATCAAACGTCAGTTCAAGTGTATCAATATCGACC CGTATGCCAATGCTTTCAACATGAATTCCGAAGGCGGCGAATGG ATGAGTGACCTTACGGACATGAAGCCCGAACTGCACGAACGCA AATGGGAAATCGACTCGCTCTGTTATCCTATCCGTCTCGCTTATC ATTACTGGAAGACGACGGGAGATGCCAGTATATTCTCCGACGA

		ATGGCTTACAGCCATCGCCAAGGTTCTGAAAACGTTTAAGGAAC AGCAACGAAAAGAAGATCCGAAAGGTCCTTATCGTTTCCAACG CAAAACGGAACGTGCACTCGATACGATGACCAATGACGGCTGG GGCAATCCTGTAAAGCCGGTCGGACTGATTGCTTCTGCTTTCCG TCCTTCGGATGATGCTACAACCTTTCCAGTTTCTCGTTCCGTCCAA CTTCTTTGCTGTAACTTCATTGCGCAAAGCTGCCGAAATTCTGAA TACGGTCAACAAGAAACCTGATTTAGCTAAAGAATGTACTACAC TGTCTAACGAAGTGGAACAGCCCTGAAAAAGTATGCGGTTTAC AATCATCCGAAATATGGCAAAATCTATGCTTTCGAAGTGGACGG TTTCGGCAATCAACTGTTAATGGATGATGCCAATGTGCCGAGTC TCATTGCCCTGCCTTATCTTGGGGATGTGAAAGTGAACGATCCT ATTTATCAGAATACCCGTAAGTTTGTATGGAGCGAAGATAATCC TTACTTCTTCAAAGGTACTGCCGGCGAAGGAATTGGCGGTCCGC ACATCGGATATGATATGATTTGGCCCATGAGTATTATGATGAAA GCATTACCCAGTCAAAACGACGCAGAAATCAAGACCTGCATCA AAATGCTGATGGATACGGATGCCGGAACAGGGTTCATGCATGA ATCTTTCCACAAGAACGACCCGAAAAACTTTACTCGTTCCCTGGT TTGCATGGCAAAATACGCTGTTTGGAGAACTAATCCTAAAACTC GTGAATGAAGGAAAGGTAGACTTACTGAATAGTATCCAATAG
24	BT2632 (codon optimized)	ATGAATATTACTAAGGCCTTTTGCCTAAGTATCGCATTATTAGG AGCCTCTAATATGCAAGCCATTACCAATAGTGACTTTGTTATTC AGCAGGACAACACAAAAATCAATAATTACCAGACAAATCGTCC AGAGACATCAAAAAGGTTGTTTCGTGTCTCAGGCAGTCGAGCAG CAAATCGCTCACATCAAGCAACTTCTGACAAACGCAAGGCTTGC CTGGATGTTTCGAGAACTGCTTTCCAAACACTTTAGACACGACGG TCCACTTCGACGGAAGGACGATACATTCGTTTATACCGGCGAT ATCCACGCTATGTGGCTAAGAGACTCCGGAGCACAGGTTTGGCC CTACGTCCAACCTGGCCAATAAAGATGCCGAGCTGAAAAAATG CTGGCTGGAGTCATTAAAAGACAATTCAAATGCATTAAACATTGA TCCTTATGCAAATGCATTCAATATGAATTCAGAAGGCGGCGAGT GGATGTCCGATTTGACAGATATGAAACCCGAGCTTCATGAGCGT AAATGGGAGATCGACAGTCTTTGCTACCCCATTAGACTGGCATA TCACTATTGGAAGACAACAGGAGACGCTTCCATATTTAGTGACG AGTGGTTAACGGCAATAGCCAAAGTCCTAAAGACATTTAAGGA GCAGCAGCGTAAAGAGGACCCAAAGGGTCCATATAGATTTCAA AGAAAGACAGAGAGAGCCTTAGATACCATGACGAACGACGGAT GGGGTAATCCTGTCAAGCCTGTAGGTCTGATTGCATCCGCCTTT AGGCCATCAGATGATGCTACGACATTTCAATTCTTAGTGCCAAG TAATTTCTTTGCCGTGACTTCTCTTAGGAAAGCTGCCGAGATACT TAACACGGTAAACAAGAAACCAGAcCTTGCCAAAGAATGCACT ACATTGTCAAATGAAGTAGAAACGGCACTAAAAAATATGCCG TCTACAATCATCCCAAATACGGCAAAATCTATGCTTTTGAAGTC GATGGCTTCGGAAACCAACTATTAATGGATGACGCTAACGTTCC CTCTCTAATAGCCCTACCTTATCTTGGCGATGTAAAAGTGAACG ACCCAATCTACCAGAATACTAGAAAGTTTGTCTGGAGTGAGGAC AATCCTTACTTCTTCAAGGGTACCGCAGGAGAAGGCATCGGCGG TCCTCATATTGGTTACGATATGATTTGGCCTATGTCTATCATGAT GAAGGCCTTCACATCTCAGAATGATGCAGAGATAAAAAACATGT ATCAAAATGTTGATGGACACTGATGCCGGCACAGGTTTTATGCA TGAGTCCTTTCACAAAAACGACCCAAAAAATTTACCAGATCCT GGTTTGCTTGGCAGAACACGTTGTTTCGGAGAGTTAATTCTAAAA

		TTGGTAAACGAAGGTAAAGTCGATTTATTGAACAGTATCCAATAG
25	BT3774	ATGAACAAAAAAGTAATGCGCGTAGCCCTCGCCCTTGCCTTAGC AGGAGGAAGCTATGCACAAGATGACACCGCGAAGAAAAAGGTG AAAGCCTATATGGTGTGCGGACGCCACCTCGACACCCAGTGGAA CTGGGACATCCAGACAACAATCAACGAATATGTCTGGAATACC ATTAGTCAGAACTTATTTCTGCTAAAGAAATATCCCGAATACGT TTTCAACTTTGAAGGGGGAGTGAAGTATGCGTGGATGAAGGAA TACTATCCCGAACAGTATGAAGAGATGAAGAAATTCATCGAGG AAGGCCGCTGGCATATCGCCGGAAGTAGCTGGGAAGCAAGTGA TGTGTTGGTTCCTTCCGTCGAAGCCTCCATCCGTAACATCATGCT CGGACAGACGTACTACCGGCAAGAGTTCGGAAAAGAAGGAACG GATATCTTCTGCGGACTGCTTCGGATTCGGATGGACGCTTCC CACCATTGCCGCACACTGCGGACTGATCGGCTTCTCTTACAGA AGCTGGACTGGCGTAATCATCCCTTCTATGGAAAGAGCAAGCAT CCGTTTACCATCGGACTCTGGAAGGGCATTGACGGCAAACAAGT AATGCTAGCCACCGGATATGACTACGGACGCAAATGGAACAAC GAAGATCTCTCGAAGAATAAAGATCTGGAAAAATTAGCCCAAC GTACTCCGCTCAATACGGTCTACCGCTATTATGGAACAGGGGAT ATCGGTGGCTCTCTACTCTGGGTTCGGTACGTTCTGTAGAACA GGGAATCAAAGGTGATGGCCCGGTAGAGGTGATCAGTGCTACC AGCGATCAGTTGTTCAAAGATTATCTGCCGTTCAACAATCACCC GGAAGTGGCGGTATTTGACGGAGAGTTATTGATGGATGTTACG GAACAGGTTGCTATACTTCGCAGGCAGCCATGAAGCTGTACAAC CGGCAAAACGAACAGTTGGGCGATGCAGCAGAAAGAGCGGCGG TCGCTGCCGAATGGTTGGGTACTGCCAGCTATCCGCAACACACG CTGACGGAGGCATGGAAACGTTTCATCTTCCATCAATTCCATGA TGACCTGACGGGAACGAGTATCCCCCGTGCCTATGAGTTCTCAT GGAACGATGAACTGATCTCTCTAAAACAATTCTCACAAGTACTG ACTTCTTCCGTCAACGCCATTGCCGGACAGATGGATACACGCGT GAAAGGAACGCCTGTCGTTCTTTATAATGCAAACGCTTTCCCGG TATCGGACTTGACAGAGATCATCCTCGAACAGCCTAAAACCCCG AAAGGCTTCACTGTATACAATGCACAAGGCAAGAAAGTCGCTTC GCAAATGATCGGTTACGAGAACGGACGTGCTCACATCCTGGTTG CAGCGTCACTGCCCGCAAACAGTTATGCAGTGTACGATGTCCGC ACCGGAGGATCTGAAAAACGATCTCTCCTTCAGCCGCCTCAGC CATCGAAAACCTCCGTCTACAAAATCACACTGGATAAAAACGGA GATATCATCTCACTGACCGACAAGCGCAACAACAAGAACTCG TAAAAGATGGAAAAGCGATTTCGCCTGGCACTCTTACCGAAAA CAAGTCGTACGCATGGCCTGCATGGGAAATCCTGAAAGAGACC ATCGACCGTGAACCTGTCTCCATCACAGACGGCGCAAAGATCAC TTAGTGGAACCGGCGCACTCCGTAAAGCACTCTGCATTGAGA AGAAGTATGGCAAATCGCTCTTCAAGCAATACATCCGCCTCTAC GAAGGCAGCCGTGCCGACCGCATAGATTTCTATAACGAAATAG ACTGGCAGTCAACAAACACACTGCTGAAAGCAGAGTTTCCTCTG AATATAGAAAATGAAAAGGCTACTTACGATCTGGGAATCGGCA GCGTGGAAAGAGGTAATAATGTACAGACCGCTTACGAAGTATA TGCGCAGCAATGGGCAGACCTGACCGATAAGAACAACAGCTAC GGTGTATCGATCCTAAATGACAGTAAATATGGCTGGGATAAACC GGATAACAACACGATCCGTCTGACTCTTCTCCATACACCGGAAA CAAAAGGAAATTACGCTTATCAGGATCACAGGACTTCGGCTTC

		CATACATTTACTTATAGCCTCACAGGACATAACGGAGCACTTGA CAAACCCGCCACCGCCATCAAAGCTGAAATTCTGAATCAGCCGA TCAAAGCCTTCAGCAGTCCGAAACATGCCGGAACACTAGGTAA AGAATTTGCTTTTGTACGTTCAAGCAACGATCAAGTCGTTATCA AAGCGCTGAAAAAAGCGGAAGTATCCGATGAATATGTAGTACG TGTATATGAAACAGGAGGCGCAGCTCCGCAACAGGCAGCCATC ACCTTCGCCGGTGAATAGAGAAGGCAGTACTTGCAGACGGTA CGGAAAAAGAGATCGGCAGTGCTGACTTCAACAAGAACCAGCT GAATGTATCCATCGCTCCCTACAGCATACAGACATTTAAAGTGA AGCTGAAGAAAAAAGCTGATCTTCAAGCTCCGGCATGCGCTTAT CTTCCTTTGGACTATGATCGCAGATGTTTCAGTTGGAATGCTTTC CGCAAAGAAGGGAACCTTGAATCGGGCAACAGCTATGCAGCAG AACTTCTCCCCGACTCCATCCTGAAAGCCGACGGCATTCTTTCC GCTTGGGAGAGAAAGAAATTGCCAATGGTTTGACTTGCAAAGG CAATGTACTTCAGTTGCCAACCGGACATTCTTACAACCGTATCT ATTCCTGGCAGCCTCTGCCGGTGAAGATGCAGTTGCTACCTTC AGCACCGGTAACAACCTCACAGGAAATCACCGTACCTTCCTATAC CGGTTTTATCGGTCAGTGGGAGCATCTGGGACATACGGAAGGCT TCCTGAAAGATGCAGAAATCGCTTATGTCGGCACTCACCGTCAT GCTTCTGACAAAGATGAGGCTTATGAGTTTACGTATATGTTCAA GTTTGGCATGGATATTCCTAAAGGAGCGACTACGGTTACTTTGC CGGATCATGCAGATATCGTATTATTTGCCGCAACGCTGGTTAAT GAGAAGTATCCGGCAGTAACCTCCGGCCTCGGAATTGTTCCGCAC AGCCTTGAAAGCAGACAATGGAGAAGAAGCGACGACTAAAACA AACCTGTTGAAACAAGCCAACTAATCAAATGTTCCGGTGAAAC CAACGAAAAAGAAGTTGCAAGATATGCCGTAGACGGTGATGTG AAGACGAAATGGTGTGATACAAGCACGGCTCCCAACTACATTG ACTTCGACTTCGGAAAGGAACAGACGATCCGTGGATGGAAGTT GGTAAATGCCGGAAATGAAGGCAGCGTCTTTATCACTCATACCT GCTTCTTACAAGGCAGAAACAGTCCGGACGAAGAATGGAAAAC GATTGATGAACTGAGTGATAACAAGAAAAACACGGTAGTTTCGC CAGTTTAAGCCGACTTCGGTACGTTACGTCAGACTGCTGGTTAC ACAATCTACACAAAACAACAGTCTGAAGGCTGCAAGAATCTAC GAGTTGGAGGTTTATTGA
26	BT3780	ATGAAATCAACCTTTTTATTCTGGTTACTACAACCATGATGACT TGTACCGCCTTGGGACAACCTTCCAACGACAAAAAGAACGTATT ACCCGACTGGGCGTTTCGGAGGCTTCGAACGACCACAGGGAGCT AATCCGGTGATATCTCCTATAGAGAACACGAAATTCTATTGTCC GATGACACAGGATTACGTTGCATGGGAATCCAATGACACTTTCA ATCCGGCTGCTACCCTGCATGACGGCAAGATTGTCGTGCTGTAT CGGGCAGAAGATAAATCCGGTGTCCGGTATCCGGTACCGTACCTC ACGTCTCGGATACGCCACTTCGAGCGACGGCATTCACTTCAAGC GGGAAAAGACCCCGGTATTTTATCCGGATAACGATACTCAAAA GAAACTGGAATGGCCGGGCGGATGCGAAGACCCGCGTATCGCC GTCACAGCAGAAGGACTGTATGTGATGACCTATACGCAATGGA ACCGCCACATTCCGCGTCTGGCAATAGCCACTTCCCGCAATCTG AAAGACTGGACAAAGCACGGTCCCGCTTTTGCCAAAGCGTATG ACGGCAAGTTCTTCAATTTAGGATGCAAGTCCGGCTCCATTCTG ACCGAAGTTGTCAATGGGAAACAGGTGATCAAGAAAATCGACG GAAAATACTTCATGTATTGGGGAGAGGAACATGTGTTTGCCGCC ACTTCCGAAGATTTAGTCAACTGGACTCCATACGTAAATACGGA

		CGGCTCGCTGAGAAAAC TGTTCACCCCGTGACGGACACTTCG ACAGCCAGCTGACGGAATGCGGTCCTCCAGCTATTTATACTCCA AAGGGAATCGTACTTCTGTATAATGGTAAAAACAGTGCAAGCA GAGGCGACAAACGCTATACCGCCAATGTTTACGCTGCCGGACA AGCCCTCTTCGACGCCAATGACCCGACCCGTTTCATCACCCGTC TCGACGAACCGTTCTTCCGCCCCGATGGATAGTTTCGAAAAGAGC GGGCAGTATGTAGACGGAACGGTGTTCATCGAAGGGATGGTTT ATTATAAGGATAAATGGTATCTGTATTATGGTTGCGCAGATTCC AAGGTGGGTATGGCTATCTACAATCCGAAGAAACCTGCTGCCGC AGATCCGCTGCCCTAA
27	BT3780 (codon optimized)	ATGAAGTCTACCTTCTATTCTAGTGACGACTACCATGATGACT TGCACCGCTCTTGGACAGCCCTCCAACGACAAAAAGAACGTCTT ACCCGACTGGGCATTTGGTGGCTTTGAACGTCCACAAGGCGCTA ATCCAGTTATTTCCCCCATAGAAAATACTAAATTTTATTGCCCTA TGACGCAGGACTACGTAGCCTGGGAATCAAACGACACCTTTAAT CCTGCCGCAACTCTGCACGATGGCAAAATCGTGGTGTGTATAG AGCCGAAGACAAATCCGGCGTCGGCATCGGACATAGGACATCA AGATTGGGATACGCCACGTCCTCTGACGGTATACATTTCAAAG AGAGAAGACCCCTGTCTTTTATCCCGACAATGATACGCAGAAAA AACTTGAATGGCCTGGCGGTTGTGAGGATCCAAGGATTGCAGTG ACGGCAGAGGGACTTTATGTTATGACTTACACCCAATGGAATAG ACATATACCTCGTCTAGCAATCGCAACCTCTAGGAACCTTAAAG ATTGGACGAAACATGGCCCCGCTTTTGCTAAAGCCTACGACGGA AAGTTTTTCAATTTAGGCTGTAAGAGTGGCAGTATTTTGACAGA AGTGGTCAATGGTAAACAGGTGATCAAGAAAATCGATGGTAAG TATTTTATGTATTGGGGTGAGGAACACGTTTTTCGCAGCTACTTCT GAAGACCTGGTGAAC TGGACACCCTACGTTAATACAGATGGAA GTCTAAGGAAGTTATTTTCACCTCGTGACGGTCACTTCGACTCCC AACTAACGGAATGTGGCCCCACCCGCCATTTATACGCCTAAGGGC ATCGTACTGCTGTATAACGGTAAAAATAGTGCCAGTAGAGGCG ATAAAAGATACACCGCTAACGTATACGCAGCCGGCCAAGCTCT ATTCGATGCTAACGACCCTACCAGGTTTATACTAGATTGGACG AGCCCTTTTTTCAGGCCAATGGATTCAATTTGAGAAATCAGGCCAG TACGTAGATGGCACGGTTTTTTATTGAGGGCATGGTTTATTACAA GGATAAATGGTATCTTTATTATGGTTGTGCTGATTCTAAAGTTGG TATGGCAATATATAATCCCAAGAAGCCAGCAGCTGCAGATCCAC TTCCCTAA
28	BT3781	ATGAATATAACCAAAAACACTTTGCCTCTGCGCAGCACTTTCCGGG CGCTGCCGGCGTGCAAGCAATGGAAAACCGCGAATTTGTGACC CAGCAAGACAATACCCGGGTCAATAATTACCAGACCAACCGTC CCGAAGCCTCCAAGCGCTTATTCGTATCGCAGGAAGTGGAACGA CAGATTGACCACATCAAGCAACTACTGACCAATGCGAAACTGG CATGGATGTTTCGAGAACTGTTTTCCGAACACACTGGACACTACC GTTCACTTCGACGGAAAAGAGGACACTTTTGTATACACCGGAGA CATCCACGCCATGTGGCTCCGCGACTCCGGTGCGCAGGTATGGC CCTATGTGCAGCTTGCCAATAAAGACCCCGAACTGAAAAAGAT GCTGGCAGGAGTCATCAACCGCCAGTTTAAATGTATCAATATCG ACCCGTACGCCAACGCCTTCAACATGAACTCCGAAGGAGGCGA ATGGATGAGCGACCTGACGGACATGAAACCGGAACCTTCACGAA CGCAAATGGGAAATCGACTCTCTGCTACCCGATCCGCCTGGC ATACCATTACTGGAAAACAACGGGCGATGCCAGCGTATTCTCCG

		ACGAATGGCTGCAGGCCATTGCAAATGTGCTGAAGACTTTCAAG GAACAGCAGCGTAAGGACGACGCGAAAGGTCCGTACAGATTCC AGCGTAAGACCGAACGCGCACTCGACACCATGACCAATGACGG TTGGGGCAATCCGGTGAAACCTGTCGGACTGATTGCTTCCGCTT TCCGCCCTTCGGATGACGCTACGACTTTCCAGTTCCTCGTTCCTT CCAACTTCTTTGCCGTTACTTCCTTGCGCAAAGCTGCCGAAATTC TGAACACCGTGAAACAGGAAACCGGCGCTGGCCAAAGAATGTAC CGCACTGGCGGATGAAGTAGAAAAAGCATTAAAGAAATATGCT GTCTGCAACCATCCGAAATACGGTAAGATTTATGCTTTCGAGGT AGATGGCTTCGGCAATCAGCTACTGATGGACGACGCCAACGTGC CGAGTCTCATCGCTTTGCCTTATCTGGGTGACGTCAAAGTGACT GATCCGATTTATCAGAATACCCGCAAGTTTGTATGGAGCGAAGA CAATCCTTACTTCTTCAAAGGCAGTGCCGGGGGAAGGTATCGGAG GTCCGCATATCGGATATGACATGATATGGCCCATGAGTATCATG ATGAAAGCCTTCACCAGCCAGAATGACGCAGAAATCAAACTT GCATCAAAATGCTGATGGATACGGATGCAGGTACCGGCTTCATG CACGAATCATTCAACAAAAACGACCCGAAAAAACTTTACCCGTGC ATGGTTTGCATGGCAGAATACGTTGTTTCGGAGAGCTGATCCTCA AACTGGTCAATGAAGGCAAAGTGGACTTATTGAACAGCATTCA GTAG
29	BT3781 (codon optimized)	ATGAATATTACGAAAACTTTGTGCTTGTGTGCAGCACTAAGTGG CGCAGCCGGAGTTTCAGGGCAATGGAGAACCGTGAGTTTGTACTC AACAGGATAATACAAGAGTCAATAACTATCAAACGAACCGTCC CGAGGCATCTAAAAGATTATTCGTAAGTCAAGAAGTGGAAGG CAGATAGACCATATAAAACAGTTATTGACCAATGCCAAATTAGC ATGGATGTTTCGAAAATTGCTTCCCCAATACTCTGGACACGACCG TACATTTTCGATGGTAAAGAAGATACATTCGTTTACACCGGAGAC ATTCACGCTATGTGGCTAAGAGACTCAGGCGCTCAGGTATGGCC ATACGTTTCAGCTAGCTAATAAGGATCCCGAGCTGAAAAAGATG CTAGCTGGTGTTATTAATCGTCAGTTTAAATGTATCAATATAGAT CCCTATGCTAACGCATTTAATATGAACTCCGAGGGCGGTGAATG GATGTCTGATCTGACAGATATGAAACCCGAACTGCACGAAAGG AAATGGGAAATTGATAGTCTGTGCTACCCAATCAGACTGGCATA TCATTATTGGAAGACTACCGGTGATGCTTCCGTATTTTCCGATGA ATGGCTACAGGCCATAGCAAATGTATTAATAAACTTTCAAAGAGC AACAGAGGAAGGACGACGCAAAGGGACCTATAGATTTCAAAG GAAGACGGAAAGAGCTTTAGACACTATGACTAACGACGGCTGG GGAAATCCAGTCAAGCCAGTGGGTCTAATCGCATCCGCATTTAG GCCCTCAGATGACGCAACTACGTTCCAGTTCCTGGTCCCTTCAA ACTTCTTCGCAGTCACGTCTTTAAGGAAAGCAGCTGAGATACTA AATACGGTGAACAGAAAGCCTGCCTTGGCTAAAGAGTGACAG CACTGGCAGATGAGGTAGAGAAAGCCTTGAAGAAATACGCAGT GTGCAATCATCCCAAGTATGGCAAGATATACGCCTTCGAAGTAG ACGGCTTTGGTAATCAACTATTGATGGATGATGCTAATGTCCCT AGTTTAATAGCACTACCTTATTTAGGCGACGTAAAAGTGACGGA CCCAATTTACCAAAAATACCAGAAAATTCGTCTGGTCCGAAGATA ATCCCTACTTTTTCAAAGGTTTCAGCAGGAGAAGGTATCGGAGGA CCCCATATTGGTTACGACATGATATGGCCCATGAGTATAATGAT GAAGGCATTTACGAGTCAGAATGACGCAGAGATCAAAACCTGC ATAAAGATGCTGATGGACACTGATGCTGGCACGGGTTTTATGCA CGAGTCTTTTAATAAAAACGATCCAAAAAATTTTACCCGTGCCT

		GGTTCGCTTGGCAGAACACCTTGTTTGGAGAGTTGATACTGAAG TTGGTAAATGAAGGTAAAGTGGATCTACTGAACTCCATTCAATA G
30	BT3782	ATGAGAAATATATGTTTTGTAGCCTGTATGTTGTTTTGCCTCGCT TCCGCTTCCGGAAAAACGGTGAAAAATCATCCTTTCGTGTCCAT TGCCGACTCTATCCTCGACAATGTTCTGAATTTATATCAGACGG AAGACGGGCTGCTTACCGAAACATATCCCGTGAATCCCGACCAG AAAATCACTTATCTGGCAGGCGGAGCACAGCAGAACGGAACCT TGAAGGCCTCCTTTTCTGTGGCCCTACTCAGGGATGATGTCCGGT TGCGTAGCCATGTACCAGGCTACCGGAGACAAGAAGTACAAGA CGATACTGGAAAAGCGCATCCTGCCGGGACTGGAACAGTACTG GGACGGAGAACGCCTTCCGGCATGCTATCAGTCGTACCCTGTCA AATACGGTCAGCATGGACGCTACTACGATGACAACATCTGGATC GCACTGGATTATTGCGACTACTACCGCCTCACAAAGAAGGCCGA CTATCTGAAAAAGGCCATTGCCCTGTACGAATACATCTACAGCG GCTGGAGTGACGAACTGGGCGGAGGAATCTTCTGGTGCGAACA GCAGAAAGAAGCGAAGCATACCTGCTCCAATGCCCGGTCAACA GTACTCGGCGTCAAGCTATACCGTCTGACGAAGGACAAAAAGT ATCTGAACAAGGCCAAGGAACTTACGCATGGACCAGAAAAACA CTTGTGTGATCCCGACGACTTCCTTTACTGGGACAATATCAACCT GAAAGGGAAAGTCTCGAAAGACAAGTACGCCTACAACAGTGGA CAAATGATTCAGGCAGGTGTATTACTGTACGAAGAGACAGGAG ACAAGGATTACTTGCGCGATGCCCAGAAGACAGCCGCGGGAAC CGATGCCTTTTTCCGTTTCGAAAGCAGATAAGAAAGACCCGTCAG TCAAGGTACACAAGGATATGTCGTGGTTTAACTGTATTCTGTTC AGAGGCTTCAAGGCGCTGGAGAAGATTGACCACAACCCGACTT ATGTCCGTGCGATGGCAGAGAACGCGCTCCACGCATGGAGAAA CTACCGGGATGCCAACGGATTACTGGGCAGAGACTGGTCAGGA CATAACGAGGAACCTTATAAATGGCTGCTCGATAATGCCTGCCT GATCGAGCTGTTTCGCTGAAATCGAGAAATAA
31	BT3782 (codon optimized)	ATGCGTAACATTTGTTTTGTGCGCTTGTATGCTGTTTTGTCTTGCA TCCGCCTCTGGAAAAACTGTCAAAAATCATCCATTTGTATCCAT TGCCGACTCCATACTAGATAACGTACTAAACCTATACCAAACAG AAGACGGCCTATTAAGTAAACATATCCTGTCAACCCTGACCAG AAAATCACCTATTTGGCAGGCGGCGCTCAGCAGAACGGAACCC TAAAGGCATCCTTTCTTTGGCCTTACTCCGGTATGATGTCCGGCT GTGTGGCCATGTACCAAGCTACCGGAGACAAAAAGTACAAAAC CATACTAGAGAAGCGTATCTTACCAGGATTAGAACAATACTGGG ATGGTGAGCGTTTGCCCGCATGTTACCAATCCTATCCCGTGAAA TACGGACAACACGGCAGGTACTATGACGACAACATTTGGATTGC ATTGGACTATTGTGATTATTACCGTCTAACAAAGAAAGCAGACT ATCTGAAAAAAGCCATTGCTCTATATGAATACATATACAGTGGC TGGAGTGATGAGTTAGGTGGCGGCATCTTTTGGTGTGAGCAGCA AAAGGAAGCCAAGCACACGTGCTCCAATGCACCCTCCACGGTCT TAGGTGTTAAGCTTTACAGGCTAACGAAGGACAAGAAATACTTA AATAAGGCTAAGGAGACTTACGCCTGGACTAGAAAGCATCTTTG CGACCCCGACGACTTTTTATATTGGGATAATATTAAGTTAAAGG GAAAAGTTTCCAAAGATAAATATGCATACAACCTCTGGCCAAATG ATCCAGGCCGGAGTACTACTATACGAAGAACTGGCGATAAAG ACTACCTTAGGGATGCCCAAAAAACGGCCGCTGGTACGGACGC CTTTTCCGTTAGTAAAGCAGACAAAAAAGATCCATCAGTCAAAG

		TTCACAAAGATATGTCTTGGTTCAACGTCATCCTATTCAGAGGTT TTAAAGCTCTAGAGAAGATTGACCACAACCCAACCTTATGTGCGT GCCATGGCAGAGAATGCACTTCACGCTTGGCGTAACTATAGAGA TGCAAACGGACTTCTGGGCAGGGACTGGAGTGGCCATAATGAA GAGCCATACAAGTGGCTACTGGATAATGCCTGTCTAATAGAATT ATTCGCAGAGATCGAGAAATAA
32	BT3783	ATGAAACTAAGAAACCTTTTATTTATCGTCTTGCAGCGATAGT CTTCTGCAACTGTGACAGAGCTATCAGCCTACTTCGCTCACCGTTGC CTCCTACAACCTGAGAAATGCCAACGGTTCCGACTCCGCCCCGTG GAGACGGATGGGGACAGCGTTATCCGGTGATTGCCCAGATGGT GCAATATCACGATTTCGATATTTTCGGCACACAGGAATGCTTCC TTCACCAACTGAAAGACATGAAAGAAGCCCTTCCCGGTTATGAC TATATCGGCGTAGGCCGCGACGACGGTAAAGACAAAGGCGAAC ACTCCGCTATCTTCTACCGCACCGACAAATTCGACATCGTAGAA AAAGGAGATTTCTGGCTGTCGGAACCTCCGGACGTGCCGAGCA AAGGCTGGGATGCCGTATTGCCTCGTATTTGCAGCTGGGGGCAC TTCAAATGCAAAGATACCGGTTTTCGAGTTTCTGTTCTTCAATCTC CACATGGACCACATCGGCAAGAAAGCCCGTGTGGAGAGCGCTT TCCTCGTACAGGAAAAGATGAAAGAGCTGGGAAGAGGCAAGAA TCTGCCGGCTATCCTGACGGGAGACTTCAACGTCGACCAGACCC ACCAGTCCTACGACGCATTTGTCAGCAAAGGCGTCCTCTGTGAT TCTTACGAGAAGTGCGACTACCGATATGCGCTCAACGGAACCTT CAACAACCTTCGATCCGAACAGTTTTACCGAAAGCCGCATCGACC ATATCTTCGTTTCACCTTCTTTCCACGTCAAGAGATACGGTGTGC TGACAGATACCTATCGGAGTGTACGGGAAAACAGTAAAAAGGA GGACGTGAGAGATTGTCCGGAAGAGATCACCATTAAGGCTTAT GAAGCACGTACACCATCCGACCATTTCCTGTAAAAGTGGAAC GGTGTGTTGACCAACGTCAGCAAAAATAA
33	BT3784	ATGAAAACACATTTTTCATTTAAACACCTGTTATTTATTGGAGGT GCGGTGTTGTACAGCATGCAAATTTCTGCCGTCAAGAATCCGGT AGACTATGTCAGCACGCTGGTAGGAACGCAGTCCAAGTTTGAGT TATCGACCGGAAATACCTATCCGGCTACGGCACTGCCGTGGGGA ATGAACTTCTGGACACCGCAAACCGGTAAAATGGGCGACGGTT GGGCATATACCTACAATGCCGACAAAATCCGGGGCTTCAAACA AACACATCAACCCAGCCCGTGGATGAACGACTACGGTCAGTTTT CCATCATGCCGATCACAGGCGGACTGGTATTCGACCAGGACCAA CGTGCCAGCTGGTTCTCGCACAAGGCGGAGGTTGCCAAACCTTA TTATTATAAGGTATATCTCGCAGACCACGACGTTACTACGGAAC TCGTTCCGACGGAGCGTGCCGCTATGTTCCGTTTCACGTATCCG GAAACCAAGAACGCTTATGTCGTTATCGACGCATTCGACAAAGG CTCTTATGTAAAGGTGATTCCGGAAGAAAACAAGATCATCGGTT ATTCTACCAAGAACAGCGGCGGAGTGCCGGAGAACTTCAAGAA TTATTTTGTTCATCCAGTTCGACAAGCCGTTTACCTTTACTTCCGG CGTGAAAGAGAACAAACATTCTCCCGAACGAAACAGAAGTTCAG GGCAACCATACCGGAGCGATCATCGGATTCGCTACCCAGAAAG GGGAGATCGTTCACGCACGTGTAGCTTCTTCTTTTATCAGTTATG AGCAGGCGGAACTGAATCTCAAAGAATTGGGCAAGGATAGTTT CGACCAGCTGGTCACTAAAGGAAAAGACATCTGGAACCGTGAA ATGAGCAAAGTAGATGTGGAAGACGATAATATCGACAATCTGC GCACTTTCTATTCTGCCTCTATCGTTCGATGCTGTTCCACGCA GCTTTTATGAAATAGACGCCAAAGGACAGGTCGTACACTACAGC

		CCTTACAACGGAAAAGTGCTACCGGGCTATATGTTTACGGATAC CGGCTTCTGGGATACGTTCCGCTGTCTGTTCCCATTCCTTGAACCT GATGTATCCGTCCATGAATCAGAAGATGCAGGAAGGACTGGTC AATGCGTACCTTGAAAGCGGATTTCCTTCCGGAATGGGCAAGTCC CGGACACCGTGACTGTATGGTCGGCAACAACCTCCGCTTCCGTAG TAGCCGACGCCTATATCAAAGGACTGCGCGGATATGACATCGA AACACTTTGGGAAGCATTGAAACATGACGCAAACGCCCATCTCC GCGGCACAGCTTCGGGCCGCCTTGCATACGACGCCTACAACAAA CTGGGTTATGTCCCAACAATATCGGTATAGGACAGAATGTTGC CCGTACGCTGGAATATGCGTACAACGACTGGACCATCTACACGC TAGGCAAGAACTGGGCAAACCGGCAAGCGAAATCGACATCTT CAAACAACGTGCACTCAACTACAAGAAGTCTACCAACCCGAAA CGCAAACCTGATGGTAGGCAAAGACGACAAAGGTGTGTTCAACC CCAAATTTGATGCAGTAGACTGGAGCGGCGAGTTCTGCGAAGGT AACAGCTGGCACTGGAGTTTCTGCGTATTCCATGATCCGCAAGG ACTGATCGACCTGATGGGAGGCAAGAAAGAATTCAACAACATG ATGGATTCCGTCTTTGTCAATCCGGGCAAACAGGGTATGGAAAG CCGTGGCATGATCCACGAAATGCGTGAAATGCAGGTAATGAAC ATGGGACAGTACGCTCACGGCAACCAGCCTATCCAGCACATGGT TTATCTTTACAACCTATTCGGGAGAACCGTGGAAGGCCCAGCATT GGGTTCTGTGAAATCATGGACAAGCTCTACACGGCAGGCCCCGA CGGATATTGCGGTGACGAAGACAACGGTCAGACTTCTGCCTGGT ATGCTTCTCGGCTTTAGGATTCTACCCCGTTTGTCCGGGAACAG ATCAGTACATTCTGGGAACTCCCCTTTTCAAGTCAGCCAAGCTG CATCTGGAAAATGGAAAAACCGTCACAATAAAAGCAAGCAACA ATAACACCGACAACCGTTATGTGAAGGATATGAAGGTAAATGG CAAGGCATTACCCCGCAATTATCTGACGCACGACCAATTACTGA AAGGAGCGAATATCCAGTATCAGATGAGTCCTACGCCGAACAA ACAGCGGGGAACGACTGAAAAAGATATTCCCTATTCCCTTTCAT TTGAATAA
34	BT3788	ATGAAAAATACACAGAAATATTTTCATATTACTTTTGAITGTTGAT ATTACTTGTTCGAAGCAATATATGGGGACAAGAAACAAAAAAG GAAATTATAGTCAAAGGTGTAGTGGAAGATGATTTAGGGCCGA TAATTGGTGCGTCAGTCGTTGCTAAAAACCAGGCAGGTGTGGGA GTAATCACAAATACTGAAGGTAAGTTTTCTTTGAAAGTGGGACC TTATGATGTATTGGTAGTGACTTTTGTGTTGTTATCAGCCATATGA GCTGCCTGTTCTGAAAATGAATGATCCCAATAATGTAACATAA AGTTATTGGAAGATGTTGGCAAAATTGATGAAGTGGTAATTACA GCCAGTGGACTTCAACAAAAGAAAACCTCTGACTGGGGCAATAA CCAATGTTGATGTAAAACAGTTGAATGCTGTAGGAAGTAGTAGT CTTTCTAATTCATTGGCTGGTGTGGTTCCCGGTATTATAGCCATG CAGCGTAGTGGTGAACCGGGTGAAAATACATCTGAATTCTGGAT TCGAGGTATTAGTACCTTTGGTGCAAAATCAGGAGCCTTAGTTC TTATCGACGGAGTAGAACGAAATTTTGATGAGATTTTGCCGCAA GACATTGAATCGTTCTCAGTACTGAAAGATGCATCAGCAACTGC AATATATGGTCAGCGCGGTGCAAATGGAGTTATCTTAATTACCA CCAAGCGTGGGGAAAAAGGTAAGGTGAAAATTAATGTAAAAGC AGGATTTGACTGGAATACTCCTGTAAAAGTGCCAGAGTATGCAA GTGGTTATGATTGGGCGCGTTTAGCCAATGAGGCTCGGTTAGGA CGCTATGATTCCCCGATTTATACTCCTGAAGAATTGGAGATAAT TAGATCAGGTTTAGATACTGATTTATATCCTAATATTGATTGGA

		GGGATTTAATGTTGAAGAGTGGTGCACCTCGCTATTATGCTAAT ATTAGTTTTTCAGGTGGTAGTGATAATGTACGTTATTATGTCTCT GGACAATATAACCAGTGAACAAGGACGTTACAAAACGTTTAGCT CTGAAAATAAGTACAATACCAATACGACTTATGAACGATATAAT TATCGTGCTAACGTAGACATGAACATAACTAAGACAACAGTACT GAAAGTTAGTGTAGGTGGATGGTTGGTGAATAGGACTACGCCTA CTAGAAGTACTGGTGACATATGGGAGGATTTTGCCAAATTTACT CCTTTGTCTACTCCTCGTAAATGGTCTACAGGACAATGGCCGAG AGTGGATGGGCAAGATACTCCTGAATATCATATGACACAAAGA GGATATCATACGAAATGGGAGAGTAAGGTGGAACTAGTGTA AGTTAGAGCAGGATCTTAAGTTTATTACGCCCGGTTTGAAGTTT GAAGGAGTATTTGCTTTTGATACTTATAATGAGAATATAATAAA ACGGGAGAAAAAAGAGGAAGTATGGGAAGCCCAAAAATATAG AGATGAAAATGGTAAATTGATTTTGAAAAGAGTGGTCAATAGA AGTCCGATGAATCAAAATAAGGAAGTTAGGGGTGATAAACGAT ACTATTTTCAGGCGTCATTAGATTATAACCGTTTATTTGCTCATG CACATCGTGTTCGGTGTTTTTTGGCATGGTATACCAAGAGGAAAAG ACGGATGTAAATTTTCGACTCCAGTGATTTGATTGGTTCTATTCTT CGTCGTAATTTGGCTTATTCCGGTCGTTTTACTTATGCTTATAAA GATAAATACCTTGCTGAATTTAACTGGGGATGTACTGGTTCAGA GAATTTTGAACATGGAAAACAATTTGGTTTCTTTCCTGCTGTTTC TGCCGGTTATGTAATTTCTGAAGAGGCTTTTATGAAAAAAGCAT TGCCATGGATAGATCAATTTAAGATCAGAGCTTCTTATGGTGAG GTAGGTAATGATGTATTGGATGGTCGTCGATTCCCTTATGTGTCT CTTATAGATACTGATGATGGAGGATCATATTCATTTGGGGAATT TGGAACAAATAGAGTGCAAGCCTACCGTATTAGAACTTTGGGG ACTCCTAATTTGACTTGGGAGATAGCTAAGAAATATGATGTAGG TGTTGACTTTTCTTTTTTTAATGGGAAAATTAGTGGTGCTTTAGA TTGGTTTTTGGATAAACGTGATGACATCTTTATGCAGCGTAAAC ATATGCCATTGACTACCGGGCTTGCTGATCAGACTCCAATGGCC AATGTTCGGAAAGATGAAGTCTTATGGATGGGAAGGAAATATAG GATTTACTCAATCTATTGGTCAGGTGAATCTCCAACCTTCGTGCCA ACTTTACTTATCAGACTACTGATATCATAGATAAGGATGAAGCA GCCAATGAGTTATGGTATAAAATGGATAAAGGCTTTCAGTTAAA TCAATCGCGTGGATTGATTGCTTTAGGATTATTTAAAGATCAAG ATGAAATAGACCGTAGTCCGAAACAGACAAGTAACAGACCTAT CCTTCCCGGTGATATTAAATACAAAGATGTAAATGGTGATGGAG TTATTAATGATGATGATATTGTGCCTTTAGGATATCGGGAAGTT CCGGGATTACAGTATGGTGTTCGGTTTAAAGTGCTAATTGGAGAAA TTGGAATTTGAGTGTACTTTTCCAAGGAACAGGTAAATGTGATT TCTTTATTGGCGGTAATGGGCCTCATGCTTCCGTAGTGAACGTT ATGGTAATATTTTACAGGCAATGGTCGATGGTAATCGTTGGATA CCCAAAGAAATATCAGGCACGACTGCTACTGAAAATCCAAATG CGGATTGGCCGCTTTTGACATATGGCAATAATGATAATAATAAT AGAAAATCAACATTTTGGTTGTATGAAAGAAAATATTTGCGATT ACGGAATGTTGAAGTCAGCTATGATTTTCCACAACTTGGACGC GTAAATTTTTTGTAAAGTAACTTACGTCTAGGCTTTGTTGGACAAA ATTTGTTGACATGGGCTCCTTTTAAAATGTGGGATCCGGAAGGG ACTAGAGAGGACGGATCTAACTATCCGATAAATAAAACATTCTC ATGTTATCTTCAAATAAGCTTTTAA
--	--	--

35	BT3791	ATGAAAATTGTGAAGTATATAGTTATCGTATCATTATTTAGTATT TCTGCGATGTAGTGATGATGATGATAAAAAAACAATGAGCGAC CTGGGAATCTTGTAGAGTTACAGGTTGATGTAAATGAGATTAAT ATTGCGCAAGGAGATACCCGTAAGTAAACATTACGTCAGGTAA TGGGGAATATGTTGCGACTTCGGCTAATGAAGAAGTAGTTGTCTG CAGAAATAGATGGAAATGTGGTGAACTAACCGCTGTTGAGGG GCATAATAATGCTCAAGGAGTTGTTTATGTTAGCGATAAGTATT TCCAACGCACTAAAATTCTAGTTAATACGGCGGCAGAATTTGAA TTGAAGTTGAATAAACTTTGTTTACGCTTTATTCTCAAGTGGA GGATCTGATGAAGCTCTCATCAAGATCTATACAGGAAATGGAG GTTATTCTCTTGAAGTGATTGATGATAAAAATTGTATTGAAGTT GATCAATCTACGCTTGAAGACACAGAATCATTATGGTGAAAGG CATTGCTCAAGGTAATGCTGAGATTAAGATTACTGACCAAAAAG GAAAAGAAGCTTTTGTGAATCTGAATGTAATTGCTCCTAAGCAA ATTACGACTGATGCTGACGAAAAGGGCGTTCTGATAAATTCTAA TCAAGGATCACAAACAAGTGAAGATTCTTACAGGTAATGGAGAA TATAAGGTTCTTGATGCTGGTGATGCAAAGATCATTCGTTTGA AGTTTATGGTAATGTGGTAACGGTGACCGGAAGAAAGGCCGGA GAGACTTCATTTACTTTGACTGATGCAAAAGGACAAGTTTCACA GACTATTCATGTAAAGATCGCTCCTGAGAAGCGTTGGTATATGA ATTTAGGAAAAGAGTATGCAGTTTGGACTCACTTTGCAGAGATG ACTGGTGAGGGACTAGAGGCTGTGAAAGTTGAACTAACGGCT TTAAACTTAAAAAAATGACTTGGGAGCTAGTTGCTCGTATCGAT GGAACATAATTGGCTACAGACCTTTATGGGTAAGGAAGGCTATTT TATTCTTCGCGGTGGTGATTGGGAAAATAATAAGGGTAGACAGA TGGAGTTGGTAGGTATAGATGATAAACTAAACTGAGAACTGG ACATGGAGCCTTTGAACTCGGAAAATGGTCTCATATTGCTTTAG TTGTAGATTGTTTCGAAAGGTAAGGATGATTACAATGAAAAATAC AAGCTTTATGTTAATGGTAAACAAGTAAAGTGGGACGATAGCC GCAAAACCGATATGGACTATTCTGAGATTGATCTTTGTGCAGGT AATGACGGGGGTAGAGTATCAATCGGAAGAGCTAGTGACAACA GATGCTTTCTTGATGGTGCTATACTCGAAGCACGTATCTGGACG GTTTGTCTGACAGAGGAACAACCTTAAGGCTAATGCATGGGAGCT TCATGAACAAAATCCCGAAGGGTTATTAGGGCGCTGGGATTTCT CGGCTGGAGCTCCGACATCTTATATTGAGGATGGTACCAATTCTG GATCATGAGTTGCTGATGCATATTTTCGAAGTATGATAGCTGGAA TGCCACAGAATTTCTATGAGCAGATTTGGGGAAGCTCCCATTG AAGTACCTTTTAAATAA
36	BT3792	ATGAAAGCAATATTCAGCTGTTGATATTGAACTTTTTGACTCT GTTTATCTTTCCGTTCTTGCAGTGATGATGATAAGTCAAAGTCTGA ATTGAATGACCCCATCAGTGGCAATATTTCTCCGGTAGGTTTCAT TTGCGGTAGAAGCTACCAATAACGAGAATGAACTTCTGGTGAA ATGGACCAATCCAGTAATCGCGACGTGGATATGGTAGAACTCT CTTACAGGGACGTGGAAGCGAGTTTGTCTCGTGCTACCGACTTC TCGCCGGGACATATCATAATAACAAGTAGAGCGTGATGTCACACA GGAATATATGTTGAAGGTTTCTTATTTTGCTACTTACGAAGTTTC TGCCGTAGCTATCAGCAAAGCCGGCAAGCGATCGGTACCCGAA AGCCGTGTGGTGATGCCTTATCATGAAAAGGTGGACGAGCCGG AACTGAACTGCCGGAAATGCTGGACCGTGCACATTCTTACATG ACTTCTGTCATTGGATATTATTTTCGGCAAGAGTTCCAGAAGCTG CTGGCGTAGTAATTATCCTTATGATGGAAGGTTATTGGGATG

		GTGATGCGTTGGTCTGGGGACAAGGCGGTGGGCTTTCGGCATT GTTGCTATGCGTGATGCAACCAAAGAGAGCGAAGTGGAGAATC TTTACGGTGCAATGGATGATATGATGTTCAAAGGAATACAGTAT TTCTGTCAGCTGGATCGTGGAATCCTGGCTTATTCCTGCTACCCG GCTGCCGGTAACGAACGTTTTTACGATGATAACGTATGGATCGG GCTCGATATGGTCGACTGGTATACGGAAACGAAAGAGATGCGT TATCTGACACAGGCAAAGGTGGTATGGCGCTACCTGATCGATCA CGGTTGGGATGAGACTTGCGGAGGAGGTGTACACTGGAGGGAG TTGAACGAACACACTACCAGCAAGCACTCTTGCTCTACCGGACC TACTGCTGTGATGGGCTGTAAGATGTATCTGGCAACTCAGGAAC AGGAATATCTCGACTGGGCGATCAAATGTTACGACTATATGCTG GATGTATTGCAAGACAAGTCCGATCATTTATTCTATGACAATGT ACGCCCCGAATAAGGATGATCCCAATCTGCCGGGTGATCTTGAAA AGAACAAGTATTCCTACAACCTCCGGACAACCATTGCAGGCGGCC TGTCTCTTATATAAGATTACCGGCGAACAGAAATATCTGGATGA AGCGTATGCGATTGCTGAAAGCTGTCATAAGAAATGGTTTATGC CCTATCGTTCCAAAGAGCTGAATCTTACTTTCAATATCCTTGCTC CGGGACACGCTTGTTCAATACGATCATGTGCCGTGGATTCTTT GAACTTTATTCTATAGACAATGACCGTAAATATATCGATGATAT CGAAAAGTCAATGATTCATGCGTGGAGCAGTAGCTGTCATCAGG GTAATAACTTGCTGAATGACGATGATCTGAGAGGGGGAACTAC CAAGACCGGTTGGGAAATACTCCATCAGGGAGCATTGGTTGAAT TGTATGCCCGGTTGGCAGTATTGGAACGTGAAAACCGATAG
37	BT3792 (codon optimized)	ATGAAAGCCATTTTAAACTTCTAATTTTAAATTTCTTAACTCTT TTCATTTTCCCATCTCTGTTCTGATGATGATAAATCCAAATCTGAA TTGAACGATCCTATTTCTGGCAATATTTCTCCCGTAGGAAGTTTT GCTGTCGAGGCTACAAACAATGAAAATGAGCTTCTTGTCAAGTG GACCAATCCCAGTAACCGTGATGTGGACATGGTAGAGCTTAGTT ACAGAGACGTCGAAGCATCTCTTTCCCGTGCAACTGACTTCAGT CCCGGACACATCATCATAACAAGTTGAAAGGGATGTAACACAAG AATATATGCTTAAGGTTCCCTATTTTGCTACCTATGAGGTCTCCG CAGTTGCAATAAGTAAGGCCGGAAAGAGGTCCGTTCCCGAAAG TAGGGTAGTCATGCCTTATCACGAGAAGGTGGATGAACCTGAGT TGAAGCTGCCCCGAGATGCTGGACAGAGCACATTCCTACATGACA TCTGTAATAGGATACTACTTTGGTAAAAGTAGTCGTTCTCTGTTG GCGTTCTAACTATCCATATGACGGTAAGGGCTACTGGGACGGAG ATGCTTTAGTGTGGGGTCAGGGAGGAGGATTAAGTGCATTTGTA GCAATGCGTGATGCTACCAAGGAATCAGAGGTAGAGAATCTAT ATGGTGCTATGGACGATATGATGTTCAAGGGTATCCAATACTTC TGTCAACTAGATAGAGGTATACTGGCATAATTCTTGTTATCCTGCC GCTGGAAATGAGAGGTTTTACGATGATAATGTTTGGATTGGTCT AGATATGGTGGACTGGTATACGGAAACCAAAGAGATGAGATAC CTTACGCAAGCAAAGGTTGTATGGCGTTATTTAATTGATCACGG ATGGGACGAGACATGCGGTGGCGGCGTACATTGGAGAGAACTG AATGAACATACTACTTCAAAACACTCATGCAGTACTGGCCCCAC TGCTGTAATGGGTTGCAAGATGTATCTTGCTACGCAGGAACAAG AATACTTGGA CTGGGCAATTAAGTGTTACGATTATATGTTGGAC GTACTACAAGATAAATCAGACCACTTGTTTTATGACAACGTCAG GCCAAATAAAGATGATCCTAATTTACCAGGCGACCTAGAGAAG AATAAGTACAGTTATAATTCCGGCCAACCTCTGCAGGCCGCTTG TTTACTATATAAAATTACGGGTGAGCAAAAGTACTTGGATGAAG

		CTTATGCAATCGCCGAAAGTTGTCACAAGAAATGGTTTATGCCA TATAGAAGTAAAGAGCTAAATCTAACTTTCAACATCCTTGCCCC CGGACATGCTTGGTTTAATACTATCATGTGCCGTGGCTTTTTCGA ACTATATTCAATAGATAATGATCGTAAATACATTGATGACATAG AAAAATCAATGATACACGCCTGGAGTTCCTCCTGCCACCAGGGA AACAATCTGTAAATGACGACGACCTGAGGGGTGGTACGACCA AGACGGGCTGGGAAATTCTTCACCAAGGAGCACTGGTCGAGTT ATACGCAAGACTGGCAGTTCTTGAGAGGGAGAACCGATAG
38	BT3858	ATGATGATGAACAGATTGAATATAAAAAGAACAGTCGGCTCCT GTTTGATGGCGATGGCGTTTTTTTCGTGTACCCATACGGATCAG ACGCCCACGAAAGACTTTGTGCGATTATGTAAATCCATATATCGG CAATATCAGCCATCTGCTGGTGCCTACTTACCCAACCGTACATC TGCCGAACCTCGATGCTCAGGGTCTATCCGGAAAGGGGAGACTAT ACATCGGACAGGGTAAACGGCCTTCCGGTCGTGGTGACCAGTCA TAGAGGCAGCTCGGCTTTTAACCTGAGTCCGGTGCAGGGAGAG GTATCCCGACCGATTGTATCTTACTCCTATGATTTGGAGAATATT ACCCCTATAGTTATTCCGTATACCTGGATGAGGCTGATATACA GGTTGAGTATGCCCCTTCACATCAGGCTGGTATTTATCATATCA GTTTTGGGACGGAAGGTGATAATGCTCTGGTGGTGAATACGAAG AACGAAAGCTGGTCGCTGAAGAAAAAGGAGTCAGTGGCTATC AGGTTATTGACAACACTCCTACCAAAATCTATCTGTATCTCGAA ACCAGTCAACTACCTTTACGCAAAGGGGTACTGGCAGATGGAA AAGTTGATATGGAAAGTAAGGAAGGCAGTGCCATCGCTTTGTAT TATGGAAGCGAGAAGAACCTGAATCTACGTTACGGAATTTCTCT TATCAGTGCCGAGCAGGCAAAGAAGAATCTGCAACGTGACATC ACCACCTATGATGTAAAGGCGGTGGCGGATGCCGGACGCAGGA TATGGAACAAGACATTGGGCAAGATTGTGATAGAAGGCGGTTC GGAAGACGAAAAAGAAATCTTCTACACTTCCCTTTATCGTACCT ACGAACGCATGATCAATCTTTCGGAGGACGGGAAATATTACAGT GCTTTCGATGGCAAGATTCATGAAGATGGCGGAGTACCTTTT TACAGATGACTGGATATGGGATACTTACCGGGCTACACATCCGT TGCGTATCTTGATAGAACCGCAGAAGGAACTCGATATGATTCGT TCATATATACGGATGGCAGAACAGTCGGACAGAAGATGGATGC CTACCTTCCCCGAGGTGACCGGAGACAGTCACCGGATGAATGGC AATCATGCAGTGGCAGTTATCTGGGATGCTTATTGCAAAGGATT GAAAGACTTTGATCTGGAGGCTGCTTATGAAGCCTGCAAAGGA GCGATTACAGAAAAACGTTGTTGCCCTGGCTGAGATGTCCGTT GACGGAGCTCGATAAGTTCTATCAGGAAAAAGGATTTTCCCTG CACTGAACCCTGGCGAAGAAGAACTTGCAAGGCTGTTCATTCTG TTCGAGAGACGACAAGCGGTTGCGGTTATGTTGGGTAACTGTTA CGATAATTGGTGTCTGGCACAGATAGCCAGAACATTAAACAAG ACCGATGACTATAAGAAGTTTATGAGGATGTCTTATACGTACCG GAATGTTTATAATGCGGAAACGGGTTTCTTTCATCCCAAGAACA AGGACGGAAAGTTTATCGAACCGTTTGACTATCGATATTCGGGA GGACAGGGGGCACGTGGCTATTATGGTGAAAACAACGGTTGGA TCTATCGTTGGGATGTGCAGCACAATCCGGCGGATTTGATTGCC TTGATGGGTGGACAGGCTTCATTTATCGAGAGATTGAATCAGAC ATTCAATGAACCGTTGGGGCGGAGCAAGTTTGATTTCTATCATC AGTTGCCCGACCATACCGGTAATGTCGGCCAGTTCTCTATGGCA AATGAACCTTGTCTGCATATTCCTTATTTGTATAACTATGCCGGT CAGCCGTGGATGACACAAAAAAGGATTCGCGTTTTGCTGAACCA

		GTGGTTCCGTAATGACTTGATGGGCGTTCCCGGTGATGAAGACG GAGGTGGAATGACTGCATTTGTGGTATTCTCCATGATGGGCTTT TATCCGGTAACTCCCGGTTCTCCAACCTTATAATATCGGCAGTCC GGTATTCCAATCCGCAAAGATGGAGGTAGGTGACGGACATTATT TTGAGATCATAGCGGAGAATTATGCGCCGGACCATAAGTACATC CAGTCGGCTACCTTGAATGGAACGCCGTGGAATAAGCCGTGGTT CAGCCATGCGGATATTCAAACGGCGGACGTCTGGTTTTGCAGA TGGGAGATAAGCCCAATAAGAAGTGGGGGATAGCTTCGGATGC CGTGCCGCCCTCTTCAGAGAGTTTGCCGGAATAA
39	BT3862	ATGAGGAAAGAACTTGTGTTTTGTGTTTATTGGCATTATTCTGTGT GCCGGCTGTAACGGTAACAAAAAGAAAATGAACGGTGAACACG ATTTGGATGCGGCAAACATTACGTTGGATGACCATACGATCAGT TTTTATTATAATTGGTATGGAAATCCGTCAGTGGATGGAGAAAT GAAGCACTGGATGCACCCGATAGCCCTTGCTCCGGGACATTCGG GAGATGTCCGTGCCATATCCGGACTTAATGATGACATCGCCTGT AATTTTTATCCGGAGCTCGGAACGTACAGCAGCAATGATCCTGA AATCATTCGGAACATATCCGGATGCATATAAAAGCGAATGTCG GTGTACTGTCTGTCACTTGGTGGGGAGAAAGCGATTATGGCAAC CAAAGTGTGTCTCTCCTGCTGGATGAGGCTGCAAAGTAGGGGC AAAGGTGTGCTTTCATATAGAGCCTTTTAATGGACGCAGCCCGC AAACGGTAAGGGAGAATATTCAATATATAGTGGATACTTATGGT GATCATCCGGCTTTTTACCGTACGCACGGCAAACCTCTTTTCTTT ATCTATGATTCTTATCTGATCAAACCTGCCGAGTGGGCGAAGTT GTTTGCTGCCGGGGGAGAGATAAGTGTGCGTAATACCAAGTAC GACGGTCTTTTTATTGGTCTGACATTGAAGGAAAGCGAGTTGCC CGACATTGAGACAGCGTGCATGGATGGCTTTTACACTTACTTTG CCGCAACAGGTTTCACAAATGCTTCTACTCCGGCCAACCTGGAAA TCCATGCAGCAATGGGCAAAGGCACATAATAAATTGTTTATTCC GAGTGTCGGTCCGGGATATATTGATACCCGGATTCTGTCCTTGA ACGGAAGTACCACCCGAGACCGTGAGAATGGAAAATATTACGA TGATATGTATAAAGCTGCCATAGAAAGCGGTGCTTCTTATATTT CGATTACGTCTTCAATGAATGGCATGAAGGAACTCAGATAGAG CCGGCTGTCTCAAAGAAGTGCGATGCTTTTGAATATTTGGATTA TAAACCATTGGCTGATGATTACTATTTGATAAGAAGTGCCTATT GGGTAGATGAATTCCGGAAAGCAAGATCTGCTTCGGAAGATGTT CAATAA
40	BT3862 (codon optimized)	ATGAGGAAAGAACTTGTGTTTTGTGTTTATTGGCATTATTCTGTGT GCCGGCTGCAATGGAAATAAAAAAAAAAATGAATGGCGAGCACG ACTTGGACGCTGCCAATATTACGCTTGATGACCATACAATCTCT TTTTATTACAATTGGTACGGTAACCCATCAGTTGACGGCGAGAT GAAGCACTGGATGCACCCCATAGCACTGGCCCCCGGTCACTCCG GAGATGTTGGTGCAATATCTGGTTTGAATGATGATATTGCATGC AACTTCTACCCTGAAC TAGGAACATACTCCTCTAACGATCCTGA AATTATTCGTAAACACATTAGAATGCATATAAAGGCTAATGTAG GCGTGCTATCTGTTACCTGGTGGGGCGAGTCCGACTATGGAAAT CAGTCCGTTAGTCTACTATTAGATGAAGCTGCCAAGGTAGGTGC CAAAGTATGCTTCCACATAGAACCATTCAACGGACGTTCCCCC AAACGGTGCGTGAGAACATCCAATACATAGTAGACACCTATGG TGACCACCCCGCCTTTTATCGTACTCACGGCAAACCTTTATTTT CATTTACGACTCTTATTTGATCAAACCCGCAGAATGGGCCAAAT TGTTTGCCGCCGGCGGTGAAATATCTGTTTCGTAATACGAAGTAT

		GATGGCTTGTTTATCGGCCTTACATTAAGAATCTGAGCTACC CGATATAGAACTGCCTGCATGGACGGATTCTACACCTACTTCG CAGCTACTGGATTACGAATGCTTCAACGCCAGCCAATTGGAAA AGTATGCAACAGTGGGCTAAAGCACACAACAACTTTTCATCCC TTCTGTTGGCCCAGGATACATAGACACAAGGATAAGGCCATGG AACGGTTCTACAACCTCGTGACAGAGAGAACGGAAAGTACTACG ATGATATGTACAAAGCTGCCATAGAGTCCGGAGCCTCTTATATA TCTATCACCTCCTTTAATGAATGGCATGAGGGCACACAAATAGA GCCTGCCGTATCCAAGAAGTGCGACGCTTTCGAGTACCTTGACT ACAAACCTTTGGCCGATGACTACTATCTAATAAGGACCGCTTAC TGGGTGGATGAATTTAGGAAAGCCAGGTCTGCCTCCGAGGATGT GCAGTAA
--	--	---

Table 3: Exemplary advantageous proteins of interest (Amino Acid)

SEQ ID NO.	Sequence Info	Amino Acid sequence
41	BT2623 Bacteroides thetaiotaomi cron mannan utilization genes	MKKVIKKYFFLALAIIMYSCNEDEKYDILERYTPETITSDEIAPVL NLQAQYMDSNSEIVLVTWMNPEDDFLSKVEISCCSANDNLLGEP VLLDAVSTKVGSYQTSLSVEERGYVKIVAINNEKGVRSARTAEIL SSQQDFVYRADCLMSSVIELFFGGRYNAWNENYPNATGPYWDG IAAVWGQGAAYSGFVTMYKVTKETNNEKLRAKYAEKEETFLN SIDIFLNNGSGRKSFA YGTIYIPNDERYYDDNVWIGIEMANLYEL TGNEVYLQHANTVWNFILEGIDDVTGGGVYWKEGAVSKHTCST APAAVMALKLYQLSKNESYLEIAKSLYSYCKDVLQDPNDYLFY DNVRLSDPSDKNSSELKVSCKDKFTYNSGQPMMLAAAMLYRITKEE QFLKDAQNIAQSIYKKWFKNYHSSILDRDIMILSDPNTWFNAV FRGFVELYKIDKNDVYVKA VKNTMEHAWQSNCRNRLTNLMSD DYAGDKKEGKWNKTQGAFFVEIFSLIGELEQLGCFQE
42	BT2629	MKTHFSFKHLLFLGGAVLYSLQSSAVKNPVDYVSTLIGTQSKFE LSTGNTYPATALPWGMNFWTPQTGKMGDWAYTYDADKIRGF KQTHQPSPWMNDYGGQFAIMPITGGLVFDQDRRASWFSHKA EVA KPYYYKVYLADHDVTTELAPTERAVMFRFTYPETKNAYVIVDA FDKGSYVKVIPEENKIIGYSTKNSGGVPENFKNYFVIQFDKPFTF VSTVFENNILPNETEAKGNHTGAVIGFATKKGEIVHARVASSFIS PEQAEINLKLKGKNSFDQLVANGREIWNREMSKIEIEDDNIDNL RTFYSCLYRSMLFPRSFYEIDAKGQVMHYSYPNGEVRPGYMFT DTGFWDTRCLFPFLNLMYPSMNQKMQEGLVNTYKESGFLPEW ASPGHRDCMVGNNSASVADAYIKGLRGYDIETLWEALKHGA NAHLRGTASGRLGYESYNQLGYVANNIGIGQNVARTLEYAYND WAIYTLGKKLGKPENEIDIYKKHALNYKNVYHPERKLMVGKDN KGVFNPNFDAVDWSGEFCEGNSWHWSFCVFHDPQGLINLMGG KKEFNAMMDSVFVISGKLGMESRGMHEMREMVMNMGMQYA HGNQPIQHVMVYLYNYSSEPWKAQYWIREIMNKLYTAGPDGYC GDEDNGQTSAWYVFSALGFYVPCPGTDEYIIGTPLFKSAKLHLE NGKTITIKADNNQLDNRYIKEMKVNGKSQTRNFLTHDQLIKGAN IQFQMSPVPNKQRGTTEKDVPSLSFE
43	BT2630	MKLKNLLLIJALVAIVFCGCQS NYQPTSITVASYNLRNANGGDSIN GNGWGQRYPVIAQIVQYHDFDIFGTQECFIHQKDMKEALPGY DYIGVGRDDGKEKGEHSAIFYRTDKFDVIEKGDFWLSETPDVPS KGWDAVLPRICSWG HFCKDTGFEFLFFNLHMDHIGKKARVES AFLVQDKMKELGKGKELPAILTGFNVNQTHQS YDAFVSKGVL

		CDSYEKAGFRYAINGTFNDFDPNSFTESRIDHIFVSPSFQVKRYG VLTDYRSIVGKGEKKQANDCPEEIDIKTYQARTPSDHFVVKVEL EFDQRQQK
44	BT2631	MRNICFVACMLFCLTSAVGKTPGNTRYLSIADSILSNVLNLYQT NDGLLTETYPVNPDPQKITYLAGGTQQNGTLKASFLWPYSGMMS GCVALYKATGNKKYKKILEKRILPGMEQYWDNSRLPACYQSY TKYGQHGRYYDDNIWIALDYCDYYQLTHKPASLEKAVALYQYI YSGWSDEIGGGIFWCEQQKEAKHTCSNAPSTVLGVKLYRLTKD AKYLEKAKETYAWTKKHLCDPTDHL YWDNINLKGKVSKEKYA YNSGQMIQAGVLLYEETGDEQYLRDAQQTAAGTDAFFRTKAD KKDPTVKVHKDMAWFNVILFRGLKALYKIDKNPAYVNAMVEN ALHAWENYRDENGLLGRDWSGHNKEQYKWLLDNACLIEFFAEI
45	BT2632	MNITKAFCLSIALLGASNMQAITNSDFVIQQDNTKINNYQTNRPE TSKRLFVSQAVEQQIAHIKQLLTNARLAWMFENCFPNTLDTTVH FDGKDDTFVYTGDIHAMWLRDSGAQVWPYVQLANKDAELKK MLAGVIKRQFKCINIDPYANAFNMNSEGGEWMSDLTDMKPELH ERKWEIDSLCYPIRLAYHYWKTTGDASIFSDEWLTAAIAKVLKTF KEQQRKEDPKGPYRFQRKTERALDTMTNDGWGNPVKPVGLIAS AFRPSDDATTQFLVPSNFFAVTSLRKA AEILNTVNKKPDLAKEC TTLSNEVETALKKYAVYNHPKYGKIYAFEVDGFGNQLLMDDAN VPSLIALPYLGDVKVNDPIYQNTKRFVWSEDNPYFFKGTAGEGI GGPHIGYDMIWPM SIMMKAFTSQND AEIKTCIKMLMDTDAGTG FMHESFHKNDPKNFTRSWFAWQNTLFGELILKL VNEGKVDLLN SIQ
46	BT3774	MNKKVIAVALALALAGGSYAQDDTAKKKVKAYMVSDAHLDT QWNWDIQTITNEYVWNTISQNLFLKKYPEYVFNFEQGGVKYAW MKEYYPEQYEEMKKFIEEGRWHIAGSSWEASDVLVPSVEASIRN IMLGQTYRQEFQKEGTDIFLPDCFGFGWTLPTIAAHCGLIGFSS QKLDWRNHPFYGKSKHPFTIGLWKGIDGKQVMLAHGYDYGRK WNNEDLSKNKDLEKLAQRTPNTVYRYYYGTGDIGGSPTLGSVR SVEQGIKGDGPVEVISATSDQLFKDYLPFNNHPELPVFDGELLM DVHGTGCYTSQAAMKLYNRQNEQLGDAAERA AVAAEWLGT SYPQHTLTEAWKR FIFHQFHDDL TGT SIPRAYEFSWNDELISLKQ FSQVLTSSVNAIAGQMDTRVKGTPVVLYNANAFVSDLTEILEQ PKTPKGFTVYNAQGKKVASQMIGYENGRAHILVAASLPANSYA VYDVRTGGSEKTISPSAASAIENS VYKITL DKN GDIISLTDKRNN KELVKDGKAIRLALFTENKSYAWPAWEILKETIDREPVSITDGA KITLVENGALRKALCIEKKYGKSLFKQYIRLYEGSRADRIDFYNE IDWQSTNTLLKAEFPLNIENEKATYDLGIGSVERGNNVQTAYEV YAQQWADLTDKNNSYGV SILNDSKYGWDKPDNNTIRLTLLHTP ETKGNAYAYQDHQDFGFHTFTYSLTGHNGALDKPATAIKAEILNQ PIKAFSSPKHAGTLGKEFAFVRSSNDQVVIKALKKAEVSDEYVV RVYETGGAAPQQAAITFAGEIEKAVLADGTEKEIGSADFNKNQL NVSIAPYSIQTFKVKLKKKADLQAPACAYLPLDYDRRCFSWNAF RKEGNFESGNSYAAELLPDSILKADGIPFRLGEKEIANGLTCKGN VLQLPTGHSYNRIYFLAASAGEDAVATFSTGNNSQEITVPSYTG IGQWEHLGHTEGFLKDAEIA YVGTHR HASDKDEAYEFTYMFKF GMDIPKGATTVTLPDHADIVLFAATLVNEKYPAVTPASELFRTA LKADNGEEATTKTNLLKQAKLIKCSGETNEKEVARYAVDGDVK TKWCDTSTAPNYIDFDFGKEQTIRGWKL VNAGNEGSVFITHTCF

		LQGRNSPDEEWKTIDELSDNKKNTVVRQFKPTSVRYVRLLVTSQSTQNNSLKAARIYELEVY
47	BT3780	MKSTFLFLVTTTMMTCTALGQPSNDKKNVLPDWAFFGGFERPQG ANPVISPIENTKFYCPMTQDYVAWESNDTFNPAATLHDGKIVVL YRAEDKSGVGIGHRTSRLGYATSSDGIHFKREKTPVFYDPNDTQ KKLEWPGGCEDPRIA VTAEGLYVMTYTQWNRHIPRLAIATSRNL KDWTKHGPAFAKAYDGKFFNLGCKSGSILTEVVNGKQVIKKID GKYFMYWGEEHVFAATSEDLVNWTPYVNTDGLRKLFSRPG HFDSQLTECGPPAIYTPKGIVLLYNGKNSASRGDKRYTANVYAA GQALFDANDPTRFITRLDEPFFRPMDSFEKSGQYVDGTVFIEGM VYYKDKWYLYYGCADSKVGMAYNPKKPAAADPLP
48	BT3781	MNITKTLCLCAALSGAAGVQAMENREFVTQQDNTRVNNYQTN RPEASKRLFVSQEVERQIDHIKQLLTNAKLAWMFENCFPNTLDT TVHFDGKEDTFVYTGDIHAMWLRDSGAQVWPYVQLANKDPEL KKMLAGVINRQFKCINIDPYANAFNMNSEGGEWMSDLTDMKPE LHERKWEIDSLCYPIRLAYHYWKT TGDAVSFSDWLQAIANVL KTFKEQQRKDDAKGPYRFQRKTERALDTMTNDGWGNPVKPVG LIASAFRPSDDATTQFLVPSNFFAVTSLRKA AEILNTVNRKPAL AKECTALADEVEKALKKYAVCNHPKYGKIYAFEVDGFGNQLL MDDANVPSLIALPYLGDVKVTDPIYQNTTRKFVWSEDNPYFFKGS AGEGIGGPHIGYDMIWPM SIMMKAFTSQND AEIKTCIKMLMDT DAGTGFMHESFNKNDPKNFTRA WFAWQNTLFGELILKLVNEGK VDLLNSIQ
49	BT3782	MRNICFVACMLFCLASASGKTVKNHPFVSIADSILDNLNLYQT EDGLLTETYPVNPQKITYLAGGAQQNGTLKASFLWPYSGMMS GCVAMYQATGDKKYKTILEKRILPGLEQYWDGERLPACYQSY VKYGQHGRYYDDNIWIALDYCDYYRLTKKADYLLKKAIALYEYI YSGWSDELGGGIFWCEQQKEAKHTCSNAPSTVLGVKLYRLTKD KKYLNKAKETYAWTRKHLCDPDDFLYWDNINLKGKVSCKDYA YNSGQMIQAGVLLYEETGDKDYLRDAQKTAAGTDAFFRSKAD KKDPSVKVHKDMSWFNVILFRGFKALEKIDHNPTYVRAMAENA LHAWRNYRDANGLLGRDWSGHNEEPYKWLLDNACLIELFAEIE K
50	BT3783	MKLRNLLFIVLAAIVFCNCQSYQPTSLTVASYNLRNANGSDSAR GDGWGQRYPVIAQMVQYHDFDIFGTQECFLHQLKDMKEALPG YDYIGVGRDDGKDKGEHSAIFYRTDKFDIVEKGDFWLSETPDVP SKGWDAVLPRICSWGHEFKCKDTGFEFLFFNLHMDHIGKKARVE SAFLVQEKMKELGRGKNLPAILTGDFNVDQTHQSYDAFVSKGV LCDSYEKCDYRYALNGTFNFDPNFSFTESRIDHIFVSPSFHVCRY GVLTDTYRSVRENSKKEDVRDCPEEITIKAYEARTPSDHFPVKVE LVFDQRQK
51	BT3784	MKTHFSFKHLLFIGGAVLYSMQISAVKNPVDYVSTLVGTQSKFE LSTGNTYPATALPWGMNFWTPQTGKMGDWAYTYNADKIRGF KQTHQPSPWMNDYGGQFSIMPITGGLVFDQDQRASWFSHKA EVA KPYYYKVYLADHDVTTEL VPTERAAMFRFTYPETKNAYVVIDA FDKGSYVKVIPEENKIIGYSTKNSGGVPENFKNYFVIQFDKPFTFT SGVKENNILPNETEVQGNHTGAIIGFATQKGEIVHARVASSFISYE QAEINLKELGKDSFDQLVTKGKDIWNREMSKVDVEDDNIDNLR TFYSCLYRSMFLPRSFYEIDAKGQVVHYSPYNGKVLPGYMFDTT GFWDTFRCLFPFLNLMYPSMNQKMQEGLVNAYLESGLPEWAS PGHRDCMVGNNASV VADAYIKGLRGYDIETLWEALKHDANA

		HLRG TASGR LAYDAY NKLGYVPNNIGIGQNVARTLEYAYNDW TIYTLGKKLGKPA SEIDIFKQRALNYKNVYHPKRKLMVGKDDK GVFNPKFDAVDWSGEFCEGNSWHWSFCVFHDPQGLIDLMGGK KEFNMMDSVFVIPGKQGMESRGM IHEMREM QVMNMGQYAH GNQPIQHMVYLYNYSGE PWKAQH WVREIMDKLYTAGPDGYCG DEDNGQTS AWYVFSALGFYPCPGTDQYILGTPLFKSAKLHLEN GKTVTIKASNNNTDNRYVKDMKVNGKA FTRNYLTHDQLLKGA NIQYQMSPTPNKQRGTT EK DIPYSLSFE
52	BT3788	MKNTQKYFILLMLILLVPSNIWGQETKKEIIVKGVVEDDLGPIIG ASVVAKNQAGVG VITNTEGKFS LKVG PYDVLV VTFVGYQPYEL PVLKMNDPNNVTIK LLEDVGKIDEVVITASGLQKKTLTGAITN VDVKQLNAV GSSSLSNSLAGVVP GIIAMQRSGE PGENTSEFWIR GISTFGAKSGALVLIDGVERN FDEILPDIESFSVLKDASATAIYG QRGANGVILITTKRGEKGKVKINVKAGFDWNTPVKVPEYASGY DWARLANEARLGRYDSPIYTPEELEIIRSGLD TDLYPNIDWRDL MLKSGAPRYYANISFSGGSDNVRYYVSGQYTSEQGRYKTFSSEN KYNTNTTYERYNYRANVDMNITKT TVLKVS VGGWLVNRTTPT RSTGDIWEDFAKFTPLSTPRK WSTGQWPRVDGQDTPEYHMTQR GYHTKWESK VETSVKLEQDLKFITPGLKFEGVFAFDTYNENIIKR EKKEEVWEAQKYRDENGKLILKR VVNRSPMNQNKEVRGDKRY YFQASLDYNRLF AHAHRVGVFGMVYQEEKTDVNF DSSDLIGSIP RRNLAYSGRFTYAYKDKYLAEFNWGCTGSENF EHGKQFGFFPA VSAGYVISEEAFMKKALPWIDQFKIRASYGEVGN DVL DGRRFPY VSLIDTDDGGSYSFGEFGTNRVQAYRIRTLGTPNLTWEIAKKYD VGVDFFSFGNKISGALDWFLDKRDDIFMQRKHMPLTTGLADQT PMANVGKMKSYGWEGNIGFTQSIGQVNLQLRANFTYQTTDIID KDEAANELWYKMDKGFQLNQSRGLIALGLFKDQDEIDRSPKQT SNRPILPGDIKYKDVNGDGVINDDDIVPLGYREVPGLQYGVGLS ANWRNWNLSVLFQGTGKCDFFIGGNGPHAFRSERYGNILQAMV DGNRWIPKEISGTTATENPNADWPLLTYGNN DNNNRKSTFWLY ERKYLRLRNVEVS YDFPQTWTRKFFVSNLRLGFVGQNL LTWAP FKMWDPEGTREDGSNYPINKTFSCYLQISF
53	BT3791	MKIVKYIVIVSLFSISACSDDDDKKNNERPGNLVELQVDVNEINI AQGDTRTVNITSGNGEYVATSANEEVVVAEIDGNVVKLTAVEG HNNAQGVVYVSDKYFQRTKILVNTAAEFELKLNKTLFTLYSQV EGSDEALIKIYTGNGGYSLEVIDDKNCIEVDQSTLEDTESFMVKG IAQGN AEIKITDQKGKEAFVNLNVIAPKQITTD ADEKGVLINSNQ GSQQVKILTGNGEYKVLDA GDAKIIRLEVYGNVVTVTGRKAGE TSFTLTDAKGQVSQTIHVKIAPEKRWYMN LGKEYAVWTHFAEM TGEGLEAVKVETNGFKLKKMTWELVARIDGTNWLQTFMGKEG YFILRGGDWENNKGRQMELVGIDDKLKLRTGHGAFELGKWSHI ALVVDCSKGDDYNEKYKLYVNGKQVKWDDSRKTDMDYSEI DLCAGNDGGRVSIGRASDNRCFLDGAILEARIWTVCRTEEQLKA NAWELHEQNPEGLLGRWDFSAGAPTSYIEDGTNSDHELLMHISK YDSWNATEFPMSRFG EAPIEVPFK
54	BT3792	MKAIFKLLILNFLTFLFFPSCSDDDKSKSELNDPISGNISPVGSFAV EATNNENELLVKWTNPSNRD VDMVELSYRDVEASLSRATDFSP GHIIIQVERDVTQEYMLKV PYFATYEVSAVAISKAGKRSVPESRV VMPYHEKVDEPELKLPEMLDRAHSYMTSVIGY YFGKSSRSCWR SNYPYDGKGYWDGDALVWGQGGGLSAFVAMRDATKESEVEN LYGAMDDMMFKGIQYFCQLDRGILAYSCYPAAGNERFYDDNV

		WIGLDMVDWYTETKEMRYLTQAKVVWRYLIDHGWDETCGGG VHWRELNEHTTSKHSCSTGPTAVMGCKMYLATQEQEYLDWAI KCYDYMLDVLQDKSDHLFYDNRPNKDDPNLPGDLEKNKYSY NSGQPLQAACLLYKITGEQKYLDEAYAIAESCHKKWFMPIYRSK ELNLTFNILAPGHAWFNTIMCRGFFELYSIDNDRKYIDDIEKSMI HAWSSSCHQGNLLNDDDLRGGTTKTGWEILHQGALVELYAR LAVLERENR
55	BT3858	MMMNRLNIKRTVGSCLMAMAFFSCTHTDQTPTKDFVDYVNPYI GNISHLLVPTYPTVHLPNSMLRVYPERGDYTS DRVNGLPVVVTS HRGSSAFNLSPVQGEVSRPIVSYSDLENITPYSYSVYLDEADIQ VEYAPSHQAGIYHISFGTEGDNALVVNTKNGKLVAAEEKGVSGY QVIDNTPTKIYLYLETSQLPLRKGV LADGKVDMESKEGSAIALY YGSEKNLNLRYGISFISAEQAKKNLQRDITTYDVKA VADAGRRI WNKTLGKIVIEGGSEDEKEIFYTSLYRTYERMINLSEDGKYYSAF DGKIHEDGGVPFYTDDWIWD TYRATHPLRILIEPQKELDMIRSYI RMAEQSDRRWMPPTFPEVTGDSHRMNGNHAVAVIWDAYCKGL KDFDLEAA YEACKGAITEKTLLPWLRCPLTELDKFYQEKGFFPA LNPGEETCKAVHSFERRQAVAVMLGNCYDNWCLAQIARTLN KTDDYKKFMRMSYTYRNVYNAETGFFHPKNKD GKFIEPFDYRY SGGQGARGYYGENNGWIYRWDVQHNPADLIALMGGQASFIER LNQTFNEPLGRSKFDFYHQLPDHTGNVGGQFSMANEPC LHIPYLY NYAGQPWMTQKRIRVLLNQWFRNDLMGVPGDEDDGGGMTAFV VFSMMGFYPVTPGSPTYNIGSPVFQSAKMEVGDGHYFEIIAENY APDHKYIQSATLNGTPWNKPWF SHADIQNGGRLVLQM GDKPN KKWGIASDAVPPSSSESLPE
56	BT3862	MRKELVFVLLALFLCAGCNGNKKKMNGEHDLDAA NITLDDHTI SFYYNWYGNPSVDGEMKHWMPHIALAPGHSGDVGAISGLNDDI ACNFYPELGTYSSNDPEIIRKHIRMHIKANVGVL SVTWGSEDY GNQSVSLLLDEAAKVGAKVCFHIEPFNGRSPQTVRENIQYIVDT YGDHPAFYRTHGKPLFFIYDSYLIKPAEWAKLFAAGGEISVRNT KYDGLFIGLTLKESELPDIETACMDG FTYTYFAATGFTNASTPAN WKSMQQWAKAHNKLFI PSVGPYIDTRIRPWNGSTTRDRENGK YYDDMYKAAIESGASYISITSFNEWHEGTQIEPAVSKKCDAFEYL DYKPLADDYYLIRTAYWVDEFKARSASEDVQ
86	Erp1	MLLTSLQLVFACCLVLPAQVTA FYYYTSGAERKCFHKELSKGTL FQATYKAQIYDDQLQNYRDAGA QDFGLIDIEETFDDNHLVVH QKGSASGDLTFLASDSGEHKICIQPEAGGWLIKAKTKIDVEFQVG SDEKLDSKGKATIDILHAKVNV LNSKIGEIRREQKL MRDREATFR DASEAVNSRAMWWIVIQLIVLAVTCGWQMKHLGKFFVKQKIL
87	Erp2	MIKSTIALPSFFIVLILALVNSVAASSSYAPVAISLPFSKECLYYD MVTEDDSLAVGYQVLTGGNFEIDFDITAPDGSVITSEKQKKYS FLLKSFGVGKYTFCSNNYGTALKKVEITLKEKTLTDEHEADV NNDDIIANNAVEEIDRNLNKITKTLN YLRAREWRNMSTVNSTES RLTWLSILIIIIAVISIAQVLLIQFLFTGRQKNYV
88	Emp24	MASFATKFVIACFLFFSASAHN VLLPAYGRRCFFEDLSKGDELSI SFQFGDRNPQSSSQLTGDFIYGPERHEVLKTVRDTSHGEITLSAP YKGHFQYCFLNENTGIETKDVT FNIHGVVYVDLDDPNTNTLDSA VRKLSKLTREVKDEQSYIVIRER THRNTAESTNDRVKWWSIFQL GVVIANSLFQIYYLRRFFEVTSLV
89	Erv25	MQVLQLWLTTLISLVVAVQGLHFDIAASTDPEQVCIRDFVTEGQ LVVADIHSDGSVGDGQKLNLFVRDSVGNEYRRKRDFAGDVRV

		AFTAPSSSTAFDVCFENQAQYRGRSLSRAIELDIESGAEARDWNKI SANEKLKPIEVELRRVEEITDEIVDELTYLKNREERLRDTNESTN RRVRNFSILVIIVLSSLGVWQVNYLKNYFKTKHII
90	Erp3	MSNLCVLFFQFFFLAQFFAEASPLTFELNKGRKECLYTLTPEIDC TISYYFAVQQGESNDFDVNYEIFAPDDKNKPIIERSGERQGEWSFI GQHKGEYAICFYGGKAHDKIVDLDFKYNCERQDDIRNERRKAR KAQRNLRDSKTDPLQDSVENSIDTIERQLHVLERNIQYYKSRNTR NHHTVCSTEHRIVMFSIYGILLIIGMSCAQIAILEFIFRESRKHN* V*
91	Erp5	MKYNIVHGICLLFAITQAVGAVHFYAKSGETKCFYEHLSRGNLL IGDLDLYVEKDGLFEEDPESSLTITVDETDFDNDHRVLNQKNSHT GDVTFTALDTGEHRFCFTPFYSKKSATLRVFIELEIGNVEALDSK KKEDMNSLKGRVGQLTQRLSSIRKEQDAIREKEAEFRNQSESAN SKIMTWSVFQLLILLGTCAFQLRYLKNFFVKQKVV

Table 4: Exemplary Surface Display Molecules

SEQ ID NO.	Sequence Info	Sequence
57	Surface display molecule	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEG DFDVAVLPFSNSTNNGLLFINTTASIAAKEEGVSLDKREAEA (alpha factor) MKKVKKYFFLALAIIMYSCNEDEKYDILERYTPETTTSDEIAPVL NLQAQYMDSNSEIVLVTWMNPEDDFLSKVEISCCSANDNLLGEP VLLDAVSTKVGSYQTSLSVEERGYVKIVAINKEGVRSEARTAEIL SSQQDFVYRADCLMSSVIELFFGGRYNAWNENYPNATGPYWDGI AAVWGQGAAYSQFVTMYKVITKETNNEKLRAKYAEKEETFLNSI DIFLNNGSGRKSFAYGTYIGPNDERYYDDNVWIGIEMANLYELT GNEVYLQHANTVWNFILEGIDDVTGGGVYWKEGAVSKHTCSTA PAAVMALKLYQLSKNESYLEIAKSLYSYCKDVLQDPNDYLFYDN VRLSDPSPDKNSSELKVSKDKFTYNSGQPMALAAAMLRYITKEEQFL KDAQNIAQSIYKKWFKNYHSSILDRDIMILSDPNTWFNAMVFRGF VELYKIDKNDVYVKA VKNTMEHAWQSNCRNRLTNLMSDDYAG DKKEGKWNKTQGA FVEIFSLIGELEQLGCFQE (codon optimized BT2623) GSNGSSGSSGSSGSSGSSGSSGSSGSS (flex-linker) EAAAREAAAAREAAAAREAAAR (alpha-helix linker) GGGGSGGGGSGGGGS (linker) QFSNSTSASSTDVTSSSSISTSSGSVTITSSSEAPESDNGTSTAAPTET STEAPTTAIPNTGTSTEAPTTAIPNTGTSTEAPTDTTTEAPTTALPT NGTSTEAPTDTTTEAPTTGLPTNGTTSAPFPPTSLPPSNTTTTIPPYN PSTDYTTDYTVVTEYTTYCPEPTTFTTNGKTYTVTEPTTLTITDCP CTIEKPTTSTTEYTVVTEYTTYCPEPTTFTTNGKTYTVTEPTTLTI TDCPCTIEKSEAPESSVPVTESKGTTKETGVTTKQTTANPSLTVS TVVPVSSSASSHSVVINSN (Mature Sed1) GANVVVPGALGLAGVAMLFL (Sed1 propeptide)
58	Tir4 from <i>Saccharomyces cerevisiae</i>	QINELNVVLDDVKTNIA DYITLSYTPNSGFSLDQMPAGIMDIAAQ LVANPSDDSYTTL YSEVDFSAVEHMLTMVPWYSSRLLPELEAMD ASLTSSSSAATSSSEVASSSIASSTSSSVAPSSSEVVSSSVASSSSEV ASSSVASTSEATSSSAVTSSSAVSSSTESVSSSVSSSSAVSSSEAVS SSPVSSSVSSSAGPASSSVAPYNSTIASSSSTAQTSISTIAPYNSTTT

		TTPASSASSVIISTRNGTTVTETDNTLVTKETTVCDYSSTSAVPAST TGYNNSTKVSTATICSTCKEGTSTATDFSTLKT TVTVCD KSATVVSQSKTTGIVEQTENGA AKA VIGMGAGALAAVAAMLL
59	Tir4 from <i>Saccharo myces cerevisia e</i> (underlined is signal peptide, which may not be utilized in design)	<u>MAYSKITLLAALAAIAYAQTQAQ</u> INELNVVLDDVKTNADYITLS YTPNSGFSLDQMPAGIMDIAAQLVANPSDDSYTTLYSEVDFSAVE HMLTMVPWYSSRLLPELEAMDASLTSSSAATSSSEVASSSIASST SSSVAPSSSEVVSSSVASSSSEVASSSVASTSEATSSSAVTSSSAVS SSTESVSSSSVSSSSAVSSSEAVSSSPVSSVVSSSAGPASSSVAPYN STIASSSSTAQTSISTIA PYNSTTTTTTPASSASSVIISTRNGTTVTETD NTLVTKETTVCDYSSTSAVPASTTGYNNSTKVSTATICSTCKEGT STATDFSTLKT TVTVCD KSATVVSQSKTTGIVEQTENGA AKA VIGMGAGALAAVAAMLL
60	Tir4 (NP_014 652.1) from <i>Saccharo myces cerevisia e</i>	QINELNVVLDDVKTNADYITLSYTPNSGFSLDQMPAGIMDIAAQ LVANPSDDSYTTLYSEVDFSAVEHMLTMVPWYSSRLLPELEAMD ASLTSSSAATSSSEVASSSIASSTSSSVAPSSSEVVSSSVAPSSSEV VSSSVAPSSSEVVSSSVASSSSEVASSSVAPSSSEVVSSSVASSSSE VASSSVAPSSSEVVSSSVAPSSSEVVSSSVASSSSEVASSSVAPSSS EVVSSSVASSTSEATSSSAVTSSSAVSSSTESVSSSSVSSSSAVSSSE AVSSSPVSSVVSSSAGPASSSVAPYNSTIASSSSTAQTSISTIA PYN TTTTTPASSASSVIISTRNGTTVTETDNTLVTKETTVCDYSSTSAVP ASTTGYNNSTKVSTATICSTCKEGTSTATDFSTLKT TVTVCD SACQAKKSATVVSQSKTTGIVEQTENGA AKA VIGMGAGALAAVAAMLL
61	Tir4 (NP_014 652.1) from <i>Saccharo myces cerevisia e</i> (underlined is signal peptide, which may not be utilized in design)	<u>MAYSKITLLAALAAIAYAQTQAQ</u> INELNVVLDDVKTNADYITLS YTPNSGFSLDQMPAGIMDIAAQLVANPSDDSYTTLYSEVDFSAVE HMLTMVPWYSSRLLPELEAMDASLTSSSAATSSSEVASSSIASST SSSVAPSSSEVVSSSVAPSSSEVVSSSVAPSSSEVVSSSVASSSSEV ASSSVAPSSSEVVSSSVASSSSEVASSSVAPSSSEVVSSSVAPSSSE VVSSSVASSSSEVASSSVAPSSSEVVSSSVASSTSEATSSSAVTSSS AVSSSTESVSSSSVSSSSAVSSSEAVSSSPVSSVVSSSAGPASSVA PYNSTIASSSSTAQTSISTIA PYNSTTTTTTPASSASSVIISTRNGTTVT ETDNTLVTKETTVCDYSSTSAVPASTTGYNNSTKVSTATICSTCK EGTSTATDFSTLKT TVTVCD SACQAKKSATVVSQSKTTGIVEQTENGA AKA VIGMGAGALAAVAAMLL
62	Dan1 from <i>Saccharo myces cerevisia e</i>	ASVTTTTLSPYDERVNLIELAVYVSDIGAHLSEYYAFQALHKTETY PPEIAKAVFAGGDFTTMLTGISGDEVTRMITGVPWYSTRLMGAIS EALANEGIATAVPASTTEASSTSTSEASSAATESSSSSSESSAETSSN AASTQATVSSSESSAASTIASSAESSVASSVASSVASSASFANTTAP VSSTSSISVTPVQNGTDSTVTKTQASTVETTITSCSNVCSTVTK

		PVSSKAQSTATSVTSSASRVIDVTTNGANKFNNGVFGAAAIAGAA ALLL
63	Dan1 from <i>Saccharo myces cerevisia e</i> (underlin ed is signal peptide, which may not be utilized in design)	<u>MSRISILAVAAALVASATAASVTTT</u> LSPYDERVNLIELAVYVSDIG AHLSEYYAFQALHKTETYPPEIAKAVFAGGDFTTMLTGISGDEV TRMITGVPWYSTRLMGAISEALANEGIATAVPASTTEASSTSTSEAS SAATESSSSSESSAETSSNAASTQATVSSESSSAASTIASSAESSVA SSVASSVASSASFANTTAPVSSTSSISVTPVVQNGTDSTVTKTQAS TVETTITSCSNNVCSTVTKPVSSKAQSTATSVTSSASRVIDVTTNG ANKFNNGVFGAAAIAGAAALLL
64	Sed1 from <i>Saccharo myces cerevisia e</i>	QFSNSTSASSTDVTSSSSISTSSGSVTITSSSEAPESDNGTSTAAPTET STEAPTTAIPNTGTSTEAPTTAIPNTGTSTEAPDTTTEAPTTALPT NGTSTEAPDTTTEAPTTGLPTNGTTSAFPPTTSLPPSNTTTTPPYN PSTDYTTDYTVVTEYTTYCPEPTTFTTNGKTYTVTEPTTLTITDCP CTIEKPTTTSTTEYTVVTEYTTYCPEPTTFTTNGKTYTVTEPTTLTI TDCPCTIEKSEAPESSVPVTESKGTTTKETGVTTKQTTANPSLTVS TVVPVSSSASSHSVVINSNGANVVVPGALGLAGVAMLFL
65	Sed1 from <i>Saccharo myces cerevisia e</i> (which may not be utilized in design)	<u>MKLSTVLLSAGLASTTLA</u> QFSNSTSASSTDVTSSSSISTSSGSVTIT SSEAPESDNGTSTAAPTETSTEAPTTAIPNTGTSTEAPTTAIPNTGT STEAPDTTTEAPTTALPTNGTSTEAPDTTTEAPTTGLPTNGTTS AFPPTTSLPPSNTTTTPPYNPSTDYTTDYTVVTEYTTYCPEPTTFTT NGKTYTVTEPTTLTITDCPCTIEKPTTTSTTEYTVVTEYTTYCPEPT TFTTNGKTYTVTEPTTLTITDCPCTIEKSEAPESSVPVTESKGTTTK ETGVTTKQTTANPSLTVSTVVPVSSSASSHSVVINSNGANVVVPG ALGLAGVAMLFL
66	Dan4 from <i>Saccharo myces cerevisia e</i>	ITATTTLSPYDERVNLIELAVYVSDIRAHIFQYYSFNRHHKTETYP SEIAAAVFDYGDFTTRLTGISGDEVTRMITGVPWYSTRLKPAISSA LSKDGITYTAIPTSTSTTTTKSSTSTPTTTITSTSTSTPTTTSTST TPTTSTSTPTTTSTSTPTTTSTSTPTTTSTSTPTTTSTSTPTTT STTSTPTTTSTSTPTTTSTPTTTSTSTTSQTSTKSTPTTSSSTPTP TTSTPTTTSTTSTAPTSTTTSTTTSTTSTISTAPTSTTSSSTFSTSSASA SSVISTTATTSTTFASLTTPATSTASTDHTTSSVSTTNAFTTSATTT TTSDTYISSSSPSQVTSSAEPTTVSEVTSSVEPTRSSQVTSSAEPTTV SEFTSSVEPTRSSQVTSSAEPTTVSEFTSSVEPTRSSQVTSSAEPTTV SEFTSSVEPTRSSQVTSSAEPTTVSEFTSSVEPTRSSQVTSSAEPTTV SEFTSSVEPIRSSQVTSSAEPTTVSEVTSSVEPIRSSQVTTEPVSSF GSTFSEITSSAEPLSFSKATTSAESISSNQITISSELIVSSVITSSSEIPS SIEVLTSAGISSVEPTSLVGPSSDESISSTESLSATSTFTSAVVSSSK AADFFTRSTVSAKSDVSGNSSTQSTTFFATPSTPLAVSSTVVTSSST DSVSPNIPFSEISSPESSTAITSTSTSFIAERTSSLYLSSSNMSSFTLS TFTVSQSIVSSFSMEPTSSVASFASSPLLVSRRSNCSDARSSNTISS GLFSTIENVRNATSTFTNLSTDEIVITSCKSSCTNEDSVLTKTQVST

		VETTITSCSGGICTTLMSPVTTINAKANTLTTTETSTVETTITTCPG GVCSTLTVPVTTITSEATTTATISCEDNEEDITSTETELLTLETTITS CSGGICTTLMSPVTTINAKANTLTTTETSTVETTITTCSGGVCSTLT VPVTTITSEATTTATISCEDNEEDVASTKTELLTMETTITSCSGGIC TTLMSPVSSFNSKATTSNNAESTIPKAIKVSCSAGACTTLTTVDAG ISMFTRTGLSITQTTVTNCSGGTCTMLTAPIATATSKVISPIPKASS ATSI AHSSASYTVSINTNGAYNFDKDNIFGTAIVAVVALLLL
67	Dan4 from Saccharo myces cerevisia e (underlin ed is signal peptide, which may not be utilized in design)	<u>MVNISIVAGIVALATSAAAITATTTLS</u> SPYDERVNLIELAVYVSDIR AHIFQYYSFERNHHKTETYPSEIAAAVFDYGDFTTRLTGISGDEVTR MITGVPWYSTR LKPAISSALSKDGIYTAIPTSTSTTTTKSSTSTPT TTITSTTSTTSTPTTSTTSTPTTSTTSTPTTSTTSTPTTSTTSTPT TTSTTSTPTTSTTSTPTTSTTSTPTTSTTSTPTTSTTPTTSTTST TSQTSTKSTPTTSSSTSTPTTSTPTTSTTSTAPTSTTSTTSTSTI STAPTSTTSSSTFSTSSASASSVISTTATTSTTFASLTTPATSTASTD HTTSSVSTTNAFTTSATTTTTSDTYISSSSPSQVTSSAEPTTVSEVTS SVEPTRSSQVTSSAEPTTVSEFTSSVEPTRSSQVTSSAEPTTVSEFTS SVEPTRSSQVTSSAEPTTVSEFTSSVEPTRSSQVTSSAEPTTVSEFTS SVEPTRSSQVTSSAEPTTVSEFTSSVEPIRSSQVTSSAEPTTVSEVTS SVEPIRSSQVTTEPVSSFGSTFSEITSSAEPLSFSKATTSAESISSNQ ITISSELIVSSVITSSSEIPSSIEVL TSSGISSSVEPTSLVGPSSDESIST ESLSATSTFTSAVVSSSKAADFFTRSTVSAKSDVSGNSSTQSTTFF ATPSTPLAVSSTVVTSSSDSVSPNIPFSEISSSPESSTAITSTSTSFIAE RTSSLYLSSSNMSSFTLSTFTVSQSIVSSFSMEPTSSVASFASSPPL VSSRSNCSDARSSNTISSGLFSTIENVRNATSTFTNLSTDEIVITSCK SSCTNEDSVLTKTQVSTVETTITSCSGGICTTLMSPVTTINAKANT LTTTETSTVETTITTCPGGVCSTLTVPVTTITSEATTTATISCEDNEE DITSTETELLTLETTITSCSGGICTTLMSPVTTINAKANTLTTTETST VETTITTCSGGVCSTLTVPVTTITSEATTTATISCEDNEEDVASTKT ELLTMETTITSCSGGICTTLMSPVSSFNSKATTSNNAESTIPKAIKV SCSAGACTTLTTVDAGISMFTRTGLSITQTTVTNCSGGTCTMLTA PIATATSKVISPIPKASSATSI AHSSASYTVSINTNGAYNFDKDNIFG TAIVAVVALLLL
68	Sag1 from Saccharo myces cerevisia e	ININDITFSNLEITPLTANKQPDQGWATTFDFSADASSIREGDEFT LSMPHVYRIKLLNSSQTATISLADGTEAFKCYVSQQAAYLYENTT FTCTAQNDLSSYNTIDGSITFSLNFSDDGSSYEYELENAKFFKSGP MLVKLGNQMSDVVNFDPA AFTENVFHSRSTGYGSFESYHLGM YCPNGYFLGGTEKIDYDSSNNNVLDLCSSVQVYSSNDFNDWWFP QSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANV NTIDHALEFQYTCLDTIANNTYATQFSTTREFIVYQGRNLGTASA KSSFISTTTTDLTSINTSAYSTGSISTVETGNRTTSEVISHVVTSTK LSPTATTSLTIAQTSIYSTDNITVGTDIHTTSEVISDVETISRETAST VVAAPTSTTGWTGAMNTYISQFTSSSFATINSTPIISSSAVFETSDA SIVNVHTENITNTAAVPSEPTFVNATRNSLNSFCSSKQPSSPSYT SSPLVSSLVSKTLLSTSTFTPSVPTSNTYIKTKNTGYFEHTALTSS VGLNSFSETAVSSQGTKIDTFLVSSLIAYPSSASGSQLSGIQQNFTS TSLMISTYEGKASIFFSAELGSIIFLLLSYLLF
69	Sag1 from Saccharo myces cerevisia	<u>MFTFLKILWLFLALASAIN</u> INDITFSNLEITPLTANKQPDQGWTA TFDFSADASSIREGDEFTLSMPHVYRIKLLNSSQTATISLADGTEA FKCYVSQQAAYLYENTTFTCTAQNDLSSYNTIDGSITFSLNFSDDG GSSYEYELENAKFFKSGPMLVKLGNQMSDVVNFDPA AFTENVFH SGRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSNNNVLDLC

	e (underlined is signal peptide, which may not be utilized in design)	SSVQVYSSNDFNDWWFPQSYNDTNADVTCFGSNLWITLDEKLY DGEMLWVNALQSLPANVNTIDHALEFQYTCLDTIANTTYATQFS TTREFIVYQGRNLGTASAKSSFISTTTTDLTSINTSAYSTGSISTVET GNRTTSEVISHVTTSTKLSPATTSLTIAQTSIYSTDSNITVGTDIH TTSEVISDVETISRETASTVVAAPTSTTGWTGAMNTYISQFTSSSF ATINSTPIISSSAVFETSDASIVNVHTENITNTAAVPSEPTFVNATR NSLNSFCSSKQPSSPSSYTSSPLVSSLSVSKTLLSTSFTPSVPTSNTY IKTKNTGYFEHTALTTSSVGLNSFSETAVSSQGTKIDTFLVSSLIAY PSSASGSQLSGIQQNFTSTSLMISTYEGKASIFFSAELGSIIFLLLSYL LF
70	Fig2 from Saccharo myces cerevisia e	QIVFYQNSSTSLPVPTLVSTSIADFHESSSTGEVQYSSSYSYVQPSI DSFTSSSFLTSFEAPTETSSSYAVSSSLITSDTFSSYSDIFDEETSSLIS TSAASSEKASSTLSSTAQPHRTSHSSSSFELPVTAPSSSSLPSSSTSLT FTSVNPSQSWTSFNSEKSSALSSSTIDFTSSEISGSTSPKSLESFDTTG TITSSYSPSPSSKNSNQTSLSPLEPLSSSSGDLILSSTIQATTNDQTS KTIPTLV DATSSLPPTLRSSSMAPTSGSDSISHNFTSPPSKTS GNYD VLTSNSIDPSLFTTTSEYSSTQLSSLNRASKSETVNFTASIASTPFGT DSATSLIDPISSVGSTASSFVGISTANFSTQGNSNYVPESTASGSSQ YQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVSTATKTVD GVITEYVTWCPLTQTKSQAIGVSSSISSVPQASSFGSSILSSNSSTL AASNVPPESTASGSSQYQDWSSSSLPLSQTTWVVINTTNTQGSVT STTSPAYVSTATKTVDGVITEYVTWCPLTQTKSQAIGISSSTISATQ TSKPSSILTLGISTLQLSDATFKGTETINTHLMTESTSITEPTYFSGT SDSFYLCTSEVNLASSLSSYPNFSSEGSTATITNSTVTFGSTSKYP STSVSNPTEASQHVSSSVNSLTDFTSNSTETIAVISNIHKTSNKDY SLTTTQLKTSGMQTLVLSTVTTTVNGAATEYTTWC PASSIA YTTS ISYKTLVLTTEVC SHSECTPTVITSVTATSSSTIPLLSTSSSTVLSSTV SEGAKNPAASEVTINTQVSATSEATSTSTQVSATSATATASESST SQVSTASETISTLGTQNFTTTGSLLPALSTEMINTTVSRKTLIIST EVC SHSKCVPTVITEVVT SKGTPSN GHSSQTLQTEAVEVTLSSHQ TVTMSTEVC SNICTPTVITSVQMRSTPFPYLTSSSTSSSSLASTKKS SLEASSEMSTFSVSTQSLPLAFTSSEKRSTTSVSQWSNTVLTNTIM SSSNVISTNEKPSSTTSPYNFSSGYSLPSSSTPSQYSLSTATTINGI KTVYTTWCPLAEKSTVAASSQSSRSVDRFVSSSKPSSSLSQTSIQY TLSTATTTISGLKTVYTTWCPLTSKSTLGATTQTSSTAKVRITSAS SATSTSISLSTSTESESSSGYLSKGVC SGTECTQDVPTQSSSPASTL AYSPSVSTSSSSSFSTTTASTLTSTHTSVPLLPSSSSSISASSPSSSTLL STSLPSPAFTSSTLPTATAVSSSTFIASSLPLSSKSSLSPVSSSILM SQFSSSSSSSSSLASPLSLISPTVDTVSVLQPTTSIATLTCTDSQCQ QEVSTICNGSN CDDVTSTATTTPPSTVTD TMTCTGSECQKTSSSC DGYSCKVSETYKSSATISACSGEGCQASATSELNSQYVTMTSVIT PSAITTTSVEVHSTESTISITTVKPVTYTSSDTNGELITITSSSQTVIP SVTTIITRTKVAITSAPKPTTTTYVEQRLSSSGIATSFVAAASSTWIT TPIVSTYAGSASKFLCSKFFMIMVMVINFI
71	Fig2 from Saccharo myces cerevisia e (underlined)	<u>MNSFASLGLIYSVNLLTRVEA</u> QIVFYQNSSTSLPVPTLVSTSIAD FHESSSTGEVQYSSSYSYVQPSIDSFTSSSFLTSFEAPTETSSSYAVS SSLITSDTFSSYSDIFDEETSSLISTSAASSEKASSTLSSTAQPHRTSH SSSSFELPVTAPSSSSLPSSSTSLTFTSVNPSQSWTSFNSEKSSALSSSTI DFTSSEISGSTSPKSLESFDTTG TITSSYSPSPSSKNSNQTSLSPLEP LSSSSGDLILSSTIQATTNDQTSKTIPTLV DATSSLPPTLRSSSMAPT SGSDSISHNFTSPPSKTS GNYDVLTSNSIDPSLFTTTSEYSSTQLSSL

	ed is signal peptide, which may not be utilized in design)	NRASKSETVNFTASIASTPFGTDSATSLIDPISSVGSTASSFVGISTA NFSTQGNSNYVPESTASGSSQYQDWSSSSLPLSQTTWVINTTNT QGSVTSTTSPAYVSTATKTVDGVITEYVTWCPLTQTKSQAIGVSS SISSVPQASSFSGSSILSSNSSTLAASNVPPESTASGSSQYQDWSS SLPLSQTTWVINTTNTQGSVTSTTSPAYVSTATKTVDGVITEYV TWCPLTQTKSQAIGISSSTISATQTSKPSSILTLGISTLQLSDATFKG TETINTHLMTESTSITEPTYFSGTSDSFYLCTSEVNLASSLSSYPNFS SSEGSTATITNSTVTFGSTSKYPSTSVSNPTEASQHVSSSVNSLTDF TSNSTETIAVISNIHKTSSNKDYSLTTLTQLKTSGMQTLVLSTVTTT VNGAATEYTTWC PASSIAYTTSISYKTLVLTTEVC SHSECTPTVIT SVTATSSTIPLLSTSSSTVLSSTVSEGAKNPAASEVTINTQVSATSE ATSTSTQVSATSATATASESSTTSQVSTASETISTLGTQNFSTTGS LFPALSTEMINTTVVSRKTLIISTEVC SHSKCVPTVITEVVT SKGTP SNGHSSQTLQTEAVEVTLSSHQVTVMSTEVCNSICTPTVITSVQM RSTPFPLYTSSTSSSSSLASTKKSSLEASSEMSTFSVSTQSLPLAFTSS EKRSTTSVSQWSNTVLNTIMSSSSNVISTNEKPSSTTSPYNFSSGY SLPSSSTPSQYSLSTATTINGIKTVYTTWCPLAEKSTVAASSQSSR SVDRFVSSSKPSSSLQTSIQYTLSTATTISGLKTVYTTWCPLTSK STLGATTQTSSTAKVRITSASSATSTSISLSTSESESSSGYLSKGVC SGTECTQDVPTQSSSPASTLAYSPSVSTSSSSSFSTTASTLTSTHTS VPLLPSSSSISASSPSSTSLSTSLPSPAFTSSTLPTATAVSSSTFIASS LPLSSKSSLSPVSSSILMSQFSSSSSSSSSLASLPSLSISPTVDTVS VLQPTTSIATLTCTDSQCQEQEVSTICNGSNCCDDVTSTATTTPSTVT DTMTCTGSECQKTSSSSCDGYSCKVSEYKSSATISACSGEGCQA SATSELNSQYVTMTSVITPSAITTTTSVEVHSTESTISITTVKPVITYS SDTNGELITITSSSQTVIPSVTTIITRTKVAITSAPKPTTTTYVEQRLS SSGIATSFVAAASSTWITTPIVSTYAGSASKFLCSKFFMIMVMVIN FI
--	---	--

Table 5: Exemplary Proteins of Interest		
Sequence Info	SEQ ID NO:	Sequence
Ovomucoid (canonical)	92	AEVDCSRFPNATDKEGKDVLVCNKDLRPICGTDGVTYTNDCLL CAYSIEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKVMVL CNRAFNPCGTDGVTYDNECLLCAHKVEQGASVDKRHDGGCR KELAAVSVDCEYPKPDCTAEDRPLCGSDNKTYGNKCNFCNA VVESNGTLTLSHFGKC*
Ovomucoid	93	AEVDCSRFPNATDMEGKDVLVCNKDLRPICGTDGVTYTNDCLL CAYSVEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKVMVL CNRAFNPCGTDGVTYDNECLLCAHKVEQGASVDKRHDGGCR KELAAVSVDCEYPKPDCTAEDRPLCGSDNKTYGNKCNFCNA VVESNGTLTLSHFGKC*
Ovomucoid G162M F167A	94	AEVDCSRFPNATDMEGKDVLVCNKDLRPICGTDGVTYTNDCLL CAYSVEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKVMVL CNRAFNPCGTDGVTYDNECLLCAHKVEQGASVDKRHDGGCR KELAAVSVDCEYPKPDCTAEDRPLCGSDNKTYMKNACNA VVESNGTLTLSHFGKC*
Ovomucoid isoform 1	95	MAMAGVFVLFVFLCGFLPDAAFGAEVDCSRFPNATDKEGKD VLVCNKDLRPICGTDGVTYTNDCLLCAYSIEFGTNISKEHDGEC

precursor full length		KETVPMNCSSYANTTSEDGKVMVLCNRAFNPCGTDGVTYDN ECLLCAHKVEQGASVDKRHDGGCRKELAAVSVDCSEYKPKDC TAEDRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
Ovomucoid [Gallus gallus]	96	MAMAGVFVLFVLCGFLPDVFGAEVDCSRFPNATDMEGKD VLVCNKDLRPICGTDGVTYTNDCLLCAYSVEFGTNISKEHDGE CKETVPMNCSSYANTTSEDGKVMVLCNRAFNPCGTDGVTYD NECLLCAHKVEQGASVDKRHDGGCRKELAAVSVDCSEYKPKD CTAEDRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
Ovomucoid isoform 2 precursor [Gallus gallus]	97	MAMAGVFVLFVLCGFLPDAAFGAEVDCSRFPNATDKEGKD VLVCNKDLRPICGTDGVTYTNDCLLCAYSIEFGTNISKEHDGEC KETVPMNCSSYANTTSEDGKVMVLCNRAFNPCGTDGVTYDN ECLLCAHKVEQGASVDKRHDGGCRKELAAVDCSEYKPKDCTA EDRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
Ovomucoid [Gallus gallus]	98	AEVDCSRFPNATDKEGKDVLCNKDLRPICGTDGVTYNNECLL CAYSIEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKVMVL CNRAFNPCGTDGVTYDNECLLCAHKVEQGASVDKRHDGECR KELAAVSVDCSEYKPKDCTAEDRPLCGSDNKTYGNKCNFCNA VVESNGTLTLSHFGKC
Ovomucoid [Numida meleagris]	99	MAMAGVFVLFVLCGFLPDAAFGVEVDCSRFPNATNEEGKD VLVCTEDLRPICGTDGVTYSNDCLLCAYNIEYGTNISKEHDGEC REAVPVDCSRYPNMTSEEGKVLILCNKAFNPVCGTDGVTYDNE CLLCAHNVEQGTSGVKKHDGECRKELAAVDCSEYKPKACTME YRPLCGSDNKTYDNKCNFCNAVVESNGTLTLSHFGKC
PREDICTED: Ovomucoid isoform X1 [Meleagris gallopavo]	100	MQTITWRQPQGDHLRSRAPAATCRAGQYLTAMAGIFVLFV ALCGFLPDAAFGVEVDCSRFPNTTNEEGKDVLCCTEDLRPICGT DGVTHSECLLCAYNIEYGTNISKEHDGECREAVPMDCSRYPNT TNEEGKVMILCNKALNPVCGTDGVTYDNECVLCAHNLEQGT VGKKHDGGCRKELAAVSVDCSEYKPKACTLEYRPLCGSDNKTY GNKCNFCNAVVESNGTLTLSHFGKC
Ovomucoid [Meleagris gallopavo]	101	VEVDCSRFPNTTNEEGKDVLCCTEDLRPICGTDGVTHSECLLC YNIEYGTNISKEHDGECREAVPMDCSRYPNTTSEEGKVMILCNK ALNPVCGTDGVTYDNECVLCAHNLEQGTSGVKKHDGECRKEL AAVSVDCSEYKPKACTLEYRPLCGSDNKTYGNKCNFCNAVVES NGTLTLSHFGKC
PREDICTED: Ovomucoid isoform X2 [Meleagris gallopavo]	102	MQTITWRQPQGDHLRSRAPAATCRAGQYLTAMAGIFVLFV ALCGFLPDAAFGVEVDCSRFPNTTNEEGKDVLCCTEDLRPICGT DGVTHSECLLCAYNIEYGTNISKEHDGECREAVPMDCSRYPNT TNEEGKVMILCNKALNPVCGTDGVTYDNECVLCAHNLEQGT VGKKHDGGCRKELAAVDCSEYKPKACTLEYRPLCGSDNKTYG NKCNCFCNAVVESNGTLTLSHFGKC
Ovomucoid [Bambusicola thoracicus]	103	EYGTNISIKHNGECKETVPMDCSRYPANMTNEEGKVMMPCDRT YNPVCGTDGVTYDNECQLCAHNVEQGTSDVKKHDGVCCKEL AAVSVDCSEYKPKPECTAEERPICGSDNKTYGNKCNFCNAV VYVQP
Ovomucoid [Callipepla squamata]	104	VDCSRFPNTTNEEGKDVLAETKELHPICGTDGVTYSNECLLCY YNIEYGTNISKEHDGECTEAVPVDCSRYPNTTSEEGKVLIPCN DFNPVCGSDGVTYENECCLCAHNVEQGTSGVKKHDGGCRKEF AAVSVDCSEYKPKDCTLEYRPLCGSDNKTYASKCNFCNAV VIWEQEKNTRHHASHSVFFISARLVC

Ovomucoid [Colinus virginianus]	105	MLPLGLREYGTNTSKEHDGECTEAVPVDCSRYPNTTSEEGKVRI LCKKDINPVC GTDGV TYDNECLLC SHSVGQGASIDKKHDGGCR KEFAAVSVDCSEY PKPACMSEYRPLCGSDNK TYV NKC NFNCA VVYVQPWLHSRCLRPPTGT SFLGSEGRETSLT SRATDLQVAGC TAISAMEATRAAALLGLVLLSSFCELSHLCSQASCDVYRLSGS RNLACPRIFQPVCGTDNVTYPNECSLCRQMLRSRAVYKKHDGR CVKVDCTGYMRATGGLGTACSQQYSPLYATNGVIYSNKCTFCS AVANGEDIDLLAVKYPEEESWISVSPTPWRMLSAGA
Ovomucoid-like isoform X2 [Anser cygnoides domesticus]	106	MSWWGIKPALERPSQE QSTSGQPVD SGSTSTTTMAGIFVLLSLV LCCFPDAAFGVEVDCSRFPNTTNEEGKEVLLCTKDLSPICGTDG VTYSNECLLCAYNIEYGTNISKDHDGECKEAVPVDCSTYPNMT NEEGKVMLVCNKMFS P VCGTDGVTYDNECMLCAHNVEQGTS VGKKYDGKCKKEVATVDCSDY PKPACTVEYMPLCGSDNK TY DNKC NFNCAV VDSNGTLT LSHFGKC
Ovomucoid-like isoform X1 [Anser cygnoides domesticus]	107	MSSQNQLHRRRRPLPGGQDLNKYYWPHCTSDRFSWLLHVTAE QFRHCVCILYLPALERPSQE QSTSGQPVD SGSTSTTTMAGIFVLL SLVLCFPDAAFGVEVDCSRFPNTTNEEGKEVLLCTKDLSPICG TDGVTYSNECLLCAYNIEYGTNISKDHDGECKEAVPVDCSTYP NMTNEEGKVMLVCNKMFS P VCGTDGVTYDNECMLCAHNVEQ GTSVGKKYDGKCKKEVATVDCSDY PKPACTVEYMPLCGSDNK TYDNKC NFNCAV VDSNGTLT LSHFGKC
Ovomucoid [Coturnix japonica]	108	VEVDCSRFPNTTNEEGKDEVVCPDELRLICGTDGVTYNHECML CFYNKEYGTNISKEQDGECEGETVPMDCSRYPNTTSEDGKVTLIC TKDFS FVCGTDGVTYDNECMLCAHNVVQGTSVGKKHDGECR KELAAVSVDCSEY PKPACPKDYRPVCGSDNK TYSNKC NFNCA VVESNGTLTLNHFGKC
Ovomucoid [Coturnix japonica]	109	MAMAGVFLLSFALCGFLPDAAFGVEVDCSRFPNTTNEEGKDE VVCPELRLICGTDGVTYNHECMLCFYNKEYGTNISKEQDGECE GETVPMDCSRYPNTTSEDGKVTLCTKDFS FVCGTDGVTYDNE CMLCAHNIVQGTSVGKKHDGECRKELAAVSVDCSEY PKPAC KDYRPVCGSDNK TYSNKC NFNCAV VVESNGTLTLNHFGKC
Ovomucoid [Anas platyrhynchos]	110	MAGVFVLLSLVLCFPDAAFGVEVDCSRFPNTTNEEGKDVLLC TKELSPVCGTDGVTYSNECLLCAYNIEYGTNISKDHDGECKE VPADCSMYPNMTNEEGKMTLLCNKMFS P VCGTDGVTYDNEC MLCAHNVEQGTSVGKKYDGKCKKEVATVDCSGY PKPACTME YMPLCGSDNK TYGNKC NFNCAV VDSNGTLT LSHFGEC
Ovomucoid, partial [Anas platyrhynchos]	111	QVDCSRFPNTTNEEGKEVLLCTKELSPVCGTDGVTYSNECLLC AYNIEYGTNISKDHDGECKEAVPADCSMYPNMTNEEGKMTLL CNKMFS P VCGTDGVTYDNECMLCAHNVEQGTSVGKKYDGKC KKEVATVSVDCSGY PKPACTMEYMPLCGSDNK TYGNKC NFNCA AVV
Ovomucoid-like [Tyto alba]	112	MTMPGAFVVLFSVLCCFPDATFGVEVDCSTYPNTTNEEGKEVL VCSKILSPICGTDGVTYSNECLLCANNIEYGTNISKYHDGECKEF VPVNC SRYPNTTNEEGKVMLICN1KDLS P VCGTDGVTYDNECL LCAHNLEPGTSVGKKYDGECKKEIATVDCSDY PKPVCSLESMP LCGSDNK TYSNKC NFNCAV VDSNETLT LSHFGKC
Ovomucoid [Balearica regulorum gibbericeps]	113	MTMAGVFVLLSFALCCFPDAAFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VVPVDCSRYPNSTNEEGKVMLCSKDLNPVCGTDGVTYDNEC VLCAHNVESGTSVGKKYDGECKKETATVDCSDY PKPACTLEY MPFCGSDSK TYSNKC NFNCAV VDSNGTLT LSHFGKC

Turkey vulture [<i>Cathartes aura</i>] OVD (native sequence) bolded is native signal sequence	114	MTTAGVFVLLSFALCSFPDAAFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE FVPVDCSRYPNTTNEEDGKVLLCNKDLSPICGTDGVTYDNECL LCARNLEPGTSVGKKYDGECKKEIATVDCSDYPKPVCSLEYMP LCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like [<i>Cuculus canorus</i>]	115	MTTAGVFVLLSFTLCSFPDAAFGVEVDCSPYPNTTNEEGKEVLV CNKILSPICGTDGVTYSNECLLCAYNLEYGTNISKDYDGECKEV APVDCSRHPNTTNEEGKVELLCNKDLNPICGTNGVTYDNECLL CARNLESGETSIGKKYDGECKKEIATVDCSDYPKPVCTLEEMPLC GSDNKTYGNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid [<i>Antrostomus carolinensis</i>]	116	MTTAVVFVLLSFALCCFPDAAFGVEVDCSTYPNSTNEEGKDV VCPKILGPICGTDGVTYSNECLLCAYNIQYGTNVSKDHDGECKE IVPVDCSRYPNTTNEEGKVVFLLCNKNFDPVCGTDGDTYDNECM LCARSLEPGTTVGKKHDGECKREIATVDCSDYPKPTCSAEDMP LCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSRFGKC
Ovomucoid [<i>Cariama cristata</i>]	117	MTMTGVFVLLSFAICCFPDAAFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VVPVDCSKYPNTTNEEGKVLLCSKDLSPVCGTDGVTYDNECL LCARNLEPGSSVGKKYDGECKKEIATIDCSDYPKPVCSLEYMPL CGSDSKTYDNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like isoform X2 [<i>Pygoscelis adeliae</i>]	118	MTTAGVFVLLSFVLCFPDAVFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VVPVNCSRYPNNTTNEEGKVVLRC SKDLSPVCGTDGVTYDNECL MCARNLEPGAVVGKNYDGECKKEIATVDCSDYPKPVCSLEYM PLCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like [<i>Nipponia nippon</i>]	119	MTTAGVFVLLSIALCCFPDAAFGVEVDCSAYSNTTSEEGKEVLS CTKILSPICGTDGVTYSNECLLCAYNIEYGTNISKDHDGECKEV VSVDCSRYPNTTNEEGKAVLLCNKDLSPVCGTDGVTYDNECLL CAHNLEPGTSVGKKYD GACKKEIATVDCSDYPKPVCTLEYLPL CGSDSKTYSNKCDFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like [<i>Phaethon lepturus</i>]	120	MTTAGVFVLLSFALCCFPDAAFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGTTYSNECLLCAYNIEYGTNVSKDHDGECKV VPVDCSKYPNTTNEEDGKVLLCNKALSPICGTDRTYDNECLM CAHNLEPGTSVGKKHDGECQKEVATVDCSDYPKPVCSLEYMP LCGSDGKTYSNKCNCFCNAVVDSSNGTLTLSHFEKC
Ovomucoid-like isoform X1 [<i>Melopsittacus undulatus</i>]	121	MTTAGVFVLLSFVLCFFPDAAFGVEVDCSTYPNTTNEEGKEV LVCAKILSPVCGTDGVTYSNECLLCAHNIENGTNVGKDHDGKC KEAVPVDCSRYPNTTDEEGKVLLCNKDVSPVCGTDGVTYDN ECLLCAHNLEAGTSVDKKNDSECKTEDTTAAVSVDCSDYPKP VCTLEYLPLCGSDNKTYSNKCRFCNAVVDSSNGTLTLSRFGKC
Ovomucoid [<i>Podiceps cristatus</i>]	122	MTTAGVFVLLSFALCCSPDAAFGVEVDCSTYPNTTNEEGKEVL ACTKILSPICGTDGVTYSNECLLCAYNMEYGTNVSKDHDGKCK EVVPVDCSRYPNTTNEEGKVLLCNKDLSPVCGTDGVTYDNEC LLCARNLEPGASVGKKYDGECKKEIATVDCSDYPKPVCSLEHM PLCGSDSKTYSNKCTFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like [<i>Fulmarus glacialis</i>]	123	MTTAGVFVLLSFALCCFPDAAFGVEVDCSTYPNTTNEEGREVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VAPVGCSRYPNNTTNEEGKVLLCNKDLSPVCGTDGVTYDNECL LCARHLEPGTSVGKKYDGECKKEIATVDCSDYPKPVCSLEYMP LCGSDSKTYSNKCNCFCNAVLDSSNGTLTLSHFGKC

Ovomucoid [Aptenodytes forsteri]	124	MTTAGVFVLLSFALCCFPDAVFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VVPVDCSRYPNTTNEEGKVVLRCNKDLSPVCGTDGVTYDNEC LMCARNLEPGAIVGKKYDGECKKEIATVDCSDYPKPVCSSLEYM PLCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid-like isoform X1 [Pygoscelis adeliae]	125	MTTAGVFVLLSFVLCFPDAVFGVEVDCSTYPNTTNEEGKEVL VCTKILSPICGTDGVTYSNECLLCAYNIEYGTNVSKDHDGECKE VVPVDCSRYPNTTNEEGKVVLRCNKDLSPVCGTDGVTYDNECL MCARNLEPGAIVGKNYDGECKKEIATVDCSDYPKPVCSSLEYM PLCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid isoform X1 [Aptenodytes forsteri]	126	MSSQNQLPSRCRPLPGSQDLNKYYQPHCTGDRFCWLFYVTVEQ FRHCICIYLLQALERPSEHQSGQPADSRNTSTMTTAGVFVLLSF ALCCFPDAVFGVEVDCSTYPNTTNEEGKEVLVCTKILSPICGTD GVTYSNECLLCAYNIEYGTNVSKDHDGECKEVVPVDCSRYPNT TNEEGKVVLRCNKDLSPVCGTDGVTYDNECLMCARNLEPGAIV VGKKYDGECKKEIATVDCSDYPKPVCSSLEYMPLCGSDSKTYSN KCNCFCNAVVDSSNGTLTLSHFGKC
Ovomucoid, partial [Antrostomus carolinensis]	127	MTTAVVFVLLSFALCCFPDAAFVFGVEVDCSTYPNSTNEEGKDV VCPKILGPICGTDGVTYSNECLLCAYNIQYGTNVSKDHDGECKE IVPVDCSRYPNTTNEEGKVVLRCNKDNFDPVCGTDGDTYDNECM LCARSLEPGTTVGKKHDGECKREIATVDCSDYPKPTCSAEDMP LCGSDSKTYSNKCNCFCNAVVDSSNGTLTLSHFGKC
rOVD as expressed in pichia secreted form 1	128	EAEAAEVDCSRFPNATDKEGKDVLCNKDLRPICGTDGVTYTN DCLLCAYSIEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKV MVLNRAFNVPVCGTDGVTYDNECLLCAHKVEQGASVDKRHD GGCRKELAAVSVDCSEYKPKPDCTAEDRPLCGSDNKTYGNKCN FCNAVVESNGTLTLSHFGKC
rOVD as expressed in pichia secreted form 2	129	EEGVSLKREAEAAEVDCSRFPNATDKEGKDVLCNKDLRPIC GTDGVTYTNDCLLCAYSIEFGTNISKEHDGECKETVPMNCSSYA NTTSEDGKVMVLNRAFNVPVCGTDGVTYDNECLLCAHKVEQG ASVDKRHDGGCRKELAAVSVDCSEYKPKPDCTAEDRPLCGSDN KTYGNKCNFCNAVVESNGTLTLSHFGKC
rOVD [gallus] coding sequence containing an alpha mating factor signal sequence (bolded) as expressed in pichia	130	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLE GDFDVAVLFPNSTNNGLLFINTTIAAIAKEEGVSLEKREAEAA EVDCSRFPNATDKEGKDVLCNKDLRPICGTDGVTYTNDCLLC AYSIEFGTNISKEHDGECKETVPMNCSSYANTTSEDGKVMVLN RAFNVPVCGTDGVTYDNECLLCAHKVEQGASVDKRHDGGCR KELAAVSVDCSEYKPKPDCTAEDRPLCGSDNKTYGNKCNFCNA VVESNGTLTLSHFGKC
Turkey vulture OVD coding sequence containing secretion signals as expressed in pichia	131	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLE GDFDVAVLFPNSTNNGLLFINTTIAAIAKEEGVSLEKREAEAV EVDCSTYPNTTNEEGKEVLVCTKILSPICGTDGVTYSNECLLCA YNIEYGTNVSKDHDGECKEFPVDCSRYPNTTNEEDGKVLLCN KDLSPICGTDGVTYDNECLLCARNLEPGTSVGKKYDGECKKEI ATVDCSDYPKPVCSSLEYMPLCGSDSKTYSNKCNCFCNAVVDSSN GTLTLSHFGKC

bolded is an alpha mating factor signal sequence		
Turkey vulture OVD in secreted form expressed in <i>Pichia</i>	132	EAEAVEVDCSTYPNTTNEEGKEVLVCTKILSPICGTDGVTYSNE CLLCAYNIEYGTNVSKDHDGECKEFVPVDCSRYPNTTNEEDGKV VLLCNKDLSPICGTDGVTYDNECLLCARNLEPGTSVGKKYDGE CKKEIATVDCSDYKPKPVCSEYMPVLCGSDSKTYSNKCNCFCNAV VDSNGTLTSLHFGKC
Humming bird OVD (native sequence) bolded is the native signal sequence	133	MTMAGVFVLLSFILCCFPDTAFGVEVDCSIYPNTTSEEGKEVLV CTETLSPICGSDGVTYNNNECQLCAYNVEYGTNVSKDHDGECKE IVPVDCSRYPNTTEEGRVVMLCNKALSPVCGTDGVTYDNECLL CARNLESGTSVGKKFDGECKKEIATVDCTDYKPKPVCSLDYMPVLC GSDSKTYSNKCNCFCNAVMDSNGLTLNLHFGKC
Humming bird OVD coding sequence as expressed in <i>Pichia</i> bolded is an alpha mating factor signal sequence	134	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSLE GDFDVAVLPFSNSTNNGLLFINTTASIAAKEEGVSLDKREAEAV EVDCSIYPNTTSEEGKEVLVCTETLSPICGSDGVTYNNNECQLCA YNVEYGTNVSKDHDGECKEIVPVDCSRYPNTTEEGRVVMLCN KALSPVCGTDGVTYDNECLLCARNLESGTSVGKKFDGECKKEI ATVDCTDYKPKPVCSDYMPVLCGSDSKTYSNKCNCFCNAVMDSN GTLTLNLHFGKC
Humming bird OVD in secreted form from <i>Pichia</i>	135	EAEAVEVDCSIYPNTTSEEGKEVLVCTETLSPICGSDGVTYNNNE CQLCAYNVEYGTNVSKDHDGECKEIVPVDCSRYPNTTEEGRVV MLCNKALSPVCGTDGVTYDNECLLCARNLESGTSVGKKFDGE CKKEIATVDCTDYKPKPVCSDYMPVLCGSDSKTYSNKCNCFCNAV MDSNGTLTLNLHFGKC
Ovalbumin related protein X	136	MFFYNTDFRMGSISAANAFCFDVFNELKVQHTNENILYSPLSII VALAMVYMGARGNTEYQMEKALHFDSIAGLGSTQTKVQKP KCGKSVNIHLLFKELSDITASKANYSLRIANRLYAEKSRPILPIY LKCVMKLYRAGLETVNFKTASDQARQLINSWVEKQTEGQIKDL LVSSSTDLDLTLVLVNAIYFKGMWKTAFNAEDTREMPPFHVTKE ESKPVQMMCMNNSFNVAATLPAEKMKILELPFASGDL SMLVLLP DEVSGLERIEKTINFEKLTEWTNPNTMEKRRVKVYLPQMKIEEK YNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMEL SEDGIEMAGSTGVIEDIKHSPELEQFRADHPFLFLIKHNPTNTIVY FGRYWSP*
Ovalbumin related protein Y	137	MDSISVTNAKFCFDVFNEMKVHHVNENILYCPLSILTALAMVY LGARGNTESQMKKVLHFDSITGAGSTTDSQCGSSEYVHNLFKE LLSEITRPNATYSLEIADKLYVDKTF SVLPEYLSCARKFYTGGE EVNFKTAAEEARQLINSWVEKETNGQIKDLLVSSSIDFGTTMVF INTIYFKGIWKIAFNTEDTREMPFSMTKEESKPVQMMCMNNSF NVATLPAEKMKILELPYASGDL SMLVLLPDEVSGLERIEKTINF DKLREWTSTNAMAKKSMKVYLPRMKIEEKYNLTSILMALGMT DLFSRSANLTGISSVDNLMISDAVHGVFMEVNEEGTEATGSTGA IGNIKHSLELEEFRADHPFLFFIRYNPTNAILFFGRYWSP*
Ovalbumin	138	MGSIGAASMEFCFDVFKEKLVHHANENIFYCPIAIMSALAMVY LGAKDSTRTQINKVVRFDKLPFGFDSIEAQCGTSVNVHSSLRDI LNQITKPNVDVYSFSLASRLYAEERYPILPEYLQCVKELYRGGLEP

		INFQTAADQARELINSWVESQTNGIIRNVLPSSVDSQTAMVLV NAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLF RVASMASEKMKILELPFASGTMSMLVLLPDEVSGLEQLESIIINFE KLTEWTSSNVMEERKIKVYLPRMKMEEKYNLT SVLMAMGITD VFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAG VDAASVSEEFRADHPFLFCIKHIATNAVLF FGRGVSP*
Chicken Ovalbumin with bolded signal sequence	139	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLE GDFDVAVLFPFSNSTNNGLLFINTTIAASIAAKEEGVSLDKREAEAG SIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAMVYLGA KDSTRTQINKVVRFDKLPGF GDSIEAQCGTSVNVHSSLRDILNQI TKPNDVYSFSLASRLYAEERYPILPEYLQCVKELYRGGLEPINFQ TAADQARELINSWVESQTNGIIRNVLPSSVDSQTAMVLVNAIV FKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLFRVAS MASEKMKILELPFASGTMSMLVLLPDEVSGLEQLESIIINFEKLTE WTSSNVMEERKIKVYLPRMKMEEKYNLT SVLMAMGITDVFSS SANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAAS VSEEFRADHPFLFCIKHIATNAVLF FGRGVSP
Chicken OVA sequence as secreted from pichia	140	EAEAGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAM VYLGA KDSTRTQINKVVRFDKLPGF GDSIEAQCGTSVNVHSSLR DILNQITKPNDVYSFSLASRLYAEERYPILPEYLQCVKELYRGGLE EPINFQTAADQARELINSWVESQTNGIIRNVLPSSVDSQTAMV LVNAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIG LFRVASMASEKMKILELPFASGTMSMLVLLPDEVSGLEQLESIIIN FEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNLT SVLMAMGI TDVFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEA GVDAASVSEEFRADHPFLFCIKHIATNAVLF FGRGVSP
Predicted Ovalbumin [Achromobacter denitrificans]	141	MRVPAQLLGLLLLWLPGARCSIGASMEFCFDVFKELKVHHA NENIFYCPIAIMSALAMVYLGA KDSTRTQINKVVRFDKLPGF GD SIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSLASRLYAEERYP ILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGII RNVLPSSVDSQTAMVLVNAIVFKGLWEKAFKDEDTQAMPFR VTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASGTMSML VLLPDEVSGLEQLESIIINFEKLTEWTSSNVMEERKIKVYLPRMK MEEKYNLT SVLMAMGITDVFSSSANLSGISSAESLKISQAVHAA HAEINEAGREVVGSAEAGVDAASVSEEFRADHPFLFCIKHIATN AVLFFGRGVSPLEIKRAAAHHHHH
OLLAS epitope-tagged ovalbumin	142	MTSGFANELGPRLMGKLTMG SIGASMEFCFDVFKELKVHHA NENIFYCPIAIMSALAMVYLGA KDSTRTQINKVVRFDKLPGF GD SIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSLASRLYAEERYP ILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGII RNVLPSSVDSQTAMVLVNAIVFKGLWEKTFKDEDTQAMPFR VTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASGTMSML VLLPDEVSGLEQLESIIINFEKLTEWTSSNVMEERKIKVYLPRMK MEEKYNLT SVLMAMGITDVFSSSANLSGISSAESLKISQAVHAA HAEINEAGREVVGSAEAGVDAASVSEEFRADHPFLFCIKHIATN AVLFFGRGVSPSR
Serpin family protein [Achromobacter denitrificans]	143	MGGRRVRWEVYISRAGYVNRQIAWRRHHRSLTMRVPAQLLGL LLLWLPGARCSIGASMEFCFDVFKELKVHHANENIFYCPIAI MSALAMVYLGA KDSTRTQINKVVRFDKLPGF GDSIEAQCGTSV NVHSSLRDILNQITKPNDVYSFSLASRLYAEERYPILPEYLQCVK ELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLPSSVD

		SQTAMVLVNAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQ MMYQIGLFRVASMASEKMKILELPPFASGTMSMLVLLPDEVSGLE EQLESIINFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNLTSV LMAMGITDVFSSANLSGSSAESLKISQAVHAAHAEINEAGRE VVGSAEAGVDAASVSEEFRADHPFLFCIKHIATNAVLFGRVCV PLEIKRAAAHHHHHH
PREDICTED: ovalbumin isoform X1 [Meleagris gallopavo]	144	MGSIGAVSMEFCFDVFKELKVHHANENIFYSPFTIISALAMVYL GAKDSTRTQINKVVRFDKLPFGGDSVEAQCGTSVNVHSSLRDIL NQITKPNDVYSFSLASRLYAEETYPILPEYLQCVKELYRGGLESI NFQTAADQARGLINSWVESQTNGMIKNVLQPSSVDSQTAMVL VNAIVFKGLWEKAFKDEDTQAIPFRVTEQESKPVQMMYQIGLF KVASMASEKMKILELPPFASGTMSMWVLLPDEVSGLEQLETTISF EKMTEWISSNIMEERRIKVYLPRMKMEEKYNLTSVLMAMGITD LFSSANLSGSSAGSLKISQAVHAAAYAEIYEAGREVIGSAEAGA DATSVSEEFVRDHPFLYCIKHNLNLSILFFGRICISP
Ovalbumin precursor [Meleagris gallopavo]	145	MGSIGAVSMEFCFDVFKELKVHHANENIFYSPFTIISALAMVYL GAKDSTRTQINKVVRFDKLPFGGDSVEAQCGTSVNVHSSLRDIL NQITKPNDVYSFSLASRLYAEETYPILPEYLQCVKELYRGGLESI NFQTAADQARGLINSWVESQTNGMIKNVLQPSSVDSQTAMVL VNAIVFKGLWEKAFKDEDTQAIPFRVTEQESKPVQMMYQIGLF KVASMASEKMKILELPPFASGTMSMWVLLPDEVSGLEQLETTISF EKMTEWISSNIMEERRIKVYLPRMKMEEKYNLTSVLMAMGITD LFSSANLSGSSAGSLKISQAAHAAAYAEIYEAGREVIGSAEAGA DATSVSEEFVRDHPFLYCIKHNLNLSILFFGRICISP
Hypothetical protein [Bambusicola thoracicus]	146	YYRVPCMVLCTAFHPYIFIVLLFALDNSEFTMGSIGAVSMEFCF DVFKELRVHHPNENIFFCPFAIMSAMAMVYLGAKDSTRTQINK VIRFDKLPFGGDSTEAQCGKSANVHSSLDILNQITKPNDVYSF SLASRLYADETYSIQSEYLQCVNELYRGGLESINFQTAADQARE LINSWVESQTNGIIRNVLPSSVDSQTAMVLVNAIVFRGLWEKA FKDEDTQTMPFRVTEQESKPVQMMYQIGSFKVASMASEKMKIL ELPLASGTMSMLVLLPDEVSGLEQLETTISFEKLTEWTSSNVME ERKIKVYLPRMKMEEKYNLTSVLMAMGITDLFRSSANLSGISLA GNLKISQAVHAAHAEINEAGRKA VSSAEAGVDATSVSEEFRAD RPFLFCIKHIATKVVFVFFGRYTSP
Egg albumin	147	MGSIGAASMEFCFDVFKELKVHHANDNMLYSPFAILSTLAMVF LGAKDSTRTQINKVVHFDKLPFGGDSIEAQCGTSVNVHSSLRDI LNQITKQNDAYSFSLASRLYAQETYTVVPEYLQCVKELYRGGLE ESVNFQTAADQARGLINAWVESQTNGIIRNLPSSVDSQTAMV LVNAIAFKGLWEKAFKAEDTQTIPFRVTEQESKPVQMMYQIGS FKVASMASEKMKILELPPFASGTMSMLVLLPDDVSGLEQLESIISF EKLTEWTSSSIMEERKVKVYLPRMKMEEKYNLTSLLMAMGITD LFSSANLSGSSVGSGLKISQAVHAAHAEINEAGRVDVVGSAEAG VDATEEFRADHPFLFCVKHIETNAILLFGRCVSP
Ovalbumin isoform X2 [Numida meleagris]	148	MASIGAVSTEFCDVYKELRVHHANENIFYSPFTIISTLAMVYL GAKDSTRTQINKVVRFDKLPFGGDSIEAQCGTSVNVHSSLRDIL NQITKPNDVYSFSLASRLYAEETYPILPEYLQCVKELYRGGLESI NFQTAADQARELINSWVESQTSGIKNVLQPSSVNSQTAMVLVN AIYFKGLWERAFKDEDTQAIPFRVTEQESKPVQMMSQIGSFKV ASVASEKVKILELPPFVSGTMSMLVLLPDEVSGLEQLESTISTEKL TEWTSSSIMEERKIKVFLPRMRMEEKYNLTSVLMAMGMTDLFS

		SSANLSGISSAESLKISQAVHAAAYAEIYEAGREVVSSAEAGVDA TSVSEEFVRDHPFLLCIKHNPTNSILFFGRCISP
Ovalbumin isoform X1 [Numida meleagris]	149	MALCKAFHPYIFIVLLFDVDNSAFTMASIGAVSTEFCDVYKEL RVHHANENIFYSPFTIISTLAMVYLGAKDSTRTQINKVVRFDKLP GFGDSIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSLASRLYAE ETYPILPEYLQCVKELYRGGLESINFQTAADQARELINSWVESQ TSGIKNVLQPSSVNSQTAMVLVNAIYFKGLWERAFKDEDQTQAI PFRVTEQESKPVQMMSQIGSFKVASVASEKVKILELPFVSGTMS MLVLLPDEVSGLEQLESTISTEKLTEWTSSSIMEERKIKVFLPRM RMEEKYNLTSVLMAMGMTDLFSSSANLSGISSAESLKISQAVH AAAYAEIYEAGREVVSSAEAGVDATSVSEEFVRDHPFLLCIKHNPT NSILFFGRCISP
PREDICTED: Ovalbumin isoform X2 [Coturnix japonica]	150	MGSIGAASMEFCFDVFKELKVHHANDNMLYSPFAILSTLAMVF LGAKDSTRTQINKVVHFDKLPFGFGDSIEAQCGTSANVHSSLRDI LNQITKQNDAYSFSLASRLYAQETYTVVPEYLQCVKELYRGG ESVNFQTAADQARGLINAWVESQTNGIIRNILQPSSVDSQTAMV LVNAIAFKGLWEKAFKAEDTQTIPFRVTEQESKPVQMMHQIGS FKVASMASEKMKILELPFASGTMSMLVLLPDDVSGLEQLESTIS FEKLTEWTSSSIMEERKVKVYLPRMKMEEKYNLTSLLMAMGIT DLFSSSANLSGISSVGSLKISQAVHAAAYAEINEAGRDRVVGSAEA GVDATEEFRADHPFLFCVKHIETNAILLFGRCVSP
PREDICTED: ovalbumin isoform X1 [Coturnix japonica]	151	MGLCTAFHPYIFIVLLFALDNSEFTMGSIGAASMEFCFDVFKEL KVHHANDNMLYSPFAILSTLAMVFLGAKDSTRTQINKVVHFDK LPFGFGDSIEAQCGTSANVHSSLRDILNQITKQNDAYSFSLASRLY AQETYTVVPEYLQCVKELYRGGLESVNFQTAADQARGLINAW VESQTNGIIRNILQPSSVDSQTAMVLVNAIAFKGLWEKAFKAED TQTIPFRVTEQESKPVQMMHQIGSFKVASMASEKMKILELPFAS GTMSMLVLLPDDVSGLEQLESTISFEKLTEWTSSSIMEERKVKV YLPRMKMEEKYNLTSLLMAMGITDLFSSSANLSGISSVGSLKIS QAVHAAAYAEINEAGRDRVVGSAEAGVDATSVSEEFVRDHPFLFCVK HIETNAILLFGRCVSP
Egg albumin	152	MGSIGAASMEFCFDVFKELKVHHANDNMLYSPFAILSTLAMVF LGAKDSTRTQINKVVHFDKLPFGFGDSIEAQCGTSANVHSSLRDI LNQITKQNDAYSFSLASRLYAQETYTVVPEYLQCVKELYRGG ESVNFQTAADQARGLINAWVESQTNGIIRNILQPSSVDSQTAMV LVNAIAFKGLWEKAFKAEDTQTIPFRVTEQESKPVQMMHQIGS FKVASMASEKMKILELPFASGTMSMLVLLPDDVSGLEQLESTIS FEKLTEWTSSSIMEERKVKVYLPRMKMEEKYNLTSLLMAMGIT DLFSSSANLSGISSVGSLKIPQAVHAAAYAEINEAGRDRVVGSAEA GVDATEEFRADHPFLFCVKHIETNAILLFGRCVSP
ovalbumin [Anas platyrhynchos]	153	MGSIGAASTEFCFDVFRELKVQHVNNENIFYSPFSIISALAMVYL ARDNTRTQIDKVVHFDKLPFGGESMEAQCCTSVSVHSSLRDILT QITKPSDNFSLSFASRLYAEETYAILPEYLQCVKELYKGGLESISF QTAADQARELINSWVESQTNGIIRNILQPSSVDSQTTMVLVNAI YFKGMWEKAFKDEDQTQAMPFRMTEQESKPVQMMYQVGSFKV AMVTSEKMKILELPFASGMMSMFVLLPDEVSGLEQLESTISFEK LTEWTSSSTMMEERRMKVYLPRMKMEEKYNLTSVFMALGMTD LFSSSANMSGISSTVSLKMSEAVHAACVEIFEAGRDRVVGSAEAG MDVTSVSEEFRADHPFLFFIKHNPTNSILFFGRWMS
PREDICTED: ovalbumin-like	154	MGSIGAASTEFCFDVFRELKVQHVNNENIFYSPLSIISALAMVYL ARDNTRTQIDQVVHFDKIPFGGESMEAQCCTSVSVHSSLRDILT

[Anser cygnoides domesticus]		EITKPSDNFSLSFASRLYAEETYTILPEYLQCVKELYKGGLESISF QTAADQARELINSWVESQTNGIKNILQPSSVDSQTTMVLVNAI YFKGMWEKAFKDEDTQTMPFRMTEQESKPVQMMYQVGSFKL ATVTSEKVKILELPFASGMMSMCVLLPDEVSGLEQLETTISFEK LTEWTSSTTMEERRMKVYLPRMKMEEKYNLTSVFMALGMTD LFSSSANMSGISSTVSLKMSEAVHAACVEIFEAGRDRVGSAG MDVTSVSEEFRADHPFLFFIKHNPSNSILFFGRWISP
PREDICTED: Ovalbumin-like [Aquila chrysaetos canadensis]	155	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLTIISALSMVYL GARENTRAQIDKVLHFDKMPGFGDTIESQCGTSVSIHTSLKDMFT QITKPSDNYSLSFASRLYAEETYPIPEYLQCVKELYKGGLETISF QTAAEQARELINSWVESQTNGMIKNILQPSSVDPQTKMVLVNAI YFKGVWEKAFKDEDTQEVPRVTEQESKPVQMMYQIGSFKVA VMASEKMKILELPYASGQLSMLVLLPDDVSGLEQLESATFEKL MAWTSSTTMEERKMKVYLPRMKIEEKYNLTSVLMALGVTDLF SSSANLSGISSAESLKISKAVHEAFVEIYEAGSEVVGSTEAGMEV TSVSEEFRADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Haliaeetus albicilla]	156	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLTIISALSMVYL GARENTRTQIDKVLHFDKMTGFGDTVESQCGTSVSIHTSLKDIFT QITKPSDNYSLSLASRLYAEETYPIPEYLQCVKELYKGGLETVS FQTAAEQARELINSWVESQTNGMIKNILQPSSVDPQTKMVLVN AIYFKGVWEKAFKDEDTQEVPRVTEQESKPVQMMYQIGSFKV AVMASEKMKILELPYASGQLSMLVLLPDDVSGLEQLESATSEK LMEWTSSTTMEERKMKVYLPRMKIEEKYNLTSVLMALGVTDL FSSSADLSGISSAESLKISKAVHEAFVEIYEAGSEVVGSTEGGME VTSVSEEFRADHPFLFLIKHKPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Haliaeetus leucocephalus]	157	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLTIISALSMVYL GARENTRTQIDKVLHFDKMTGFGDTVESQCGTSVSIHTSLKDIFT QITKPSDNYSLSLASRLYAEETYPIPEYLQCVKELYKGGLETVS FQTAAEQARELINSWVESQTNGMIKNILQPSSVDPQTKMVLVN AIYFKGVWEKAFKDEDTQEVPRVTEQESKPVQMMYQIGSFKV AVMASEKMKILELPYASGQLSMLVLLPDDVSGLEQLESATSEK LMEWTSSTTMEERKMKVYLPRMKIEEKYNLTSVLMALGVTDL FSSSADLSGISSAESLKISKAVHEAFVEIYEAGSEVVGSTEGGME VTSFSEEFRADHPFLFLIKHKPTNSILFFGRCFSP
PREDICTED: Ovalbumin [Fulmarus glacialis]	158	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYL GARENTRAQIDKVVHFDKITGFGETIESQCGTSVSVHTSLKDMFT QITKPSDNYSLSFASRLYAEETYPIPEYLQCVKELYKGGLETTS FQTAAEQARELINSWVESQTNGMIKNILQPGSVPDQTEMVLVN AIYFKGMWEKAFKDEDTQAVPFRMTEQESKTVQMMYQIGSFK VAVMASEKMKILELPYASGELSMLVMLPDDVSGLEQLETAITF EKLMEWTSSNMMEERKMKVYLPRMKMEEKYNLTSVLMALG VTDLFSSSANLSGISSAESLKMKSEAVHEAFVEIYEAGSEVVGSTG AGMEVTSVSEEFRADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Chlamydotis macqueenii]	159	MGSIGAASTEFCFDVFKELRVQHVNNENVCYSPLIISALSLVYL GARENTRAQIDKVVHFDKITGFGESIESQCGTSVSVHTSLKDMFN QITKPSDNYSLSVASRLYAEERYPIPEYLQCVKELYKGGLESIS FQTAAEQAREAINSWVESQTNGMIKNILQPSSVDPQTEMVLVN AIYFKGMWQKAFKDEDTQAVPFRISEQESKPVQMMYQIGSFKV AVMAAEKMKILELPYASGELSMLVLLPDEVSGLEQLENAITVE KLMEWTSSSPMEERIMKVYLPRMKIEEKYNLTSVLMALGITDL

		FSSSANLSGISAEESLKMSEAVHQAF AEISEAGSEVVGSSSEAGID ATSVSEEFRADHPFLFLIKHNATNSILFFGRFCFSP
PREDICTED: Ovalbumin like [Nipponia nippon]	160	MGSISAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIEKVVFHDKITGFGEIESQCSTSVSVHTSLKDMFTQ ITKPSDNYSLSFASRFYAEETYPILPEYLQCVKELYKGGLETINF RTAADQARELINSWVESQTNNGMIKNILQPGSVDPQTDMLVNA IYFKGMWEKAFKDEDTQALPFRVTEQESKPVQMMYQIGSFKV AVLASEKVKILELPYASGQLSMLVLLPDDVSGLEQLETAITVEK LMEWTSSNNMEERKIKVYLPRIKIEEKYNLTSVLMALGITDLFS SSANLSGISSAESLKVSEAIHEAFVEIYEAGSEVAGSTEAGIEVTS VSEEFRADHPFLFLIKHNATNSILFFGRFCFSP
PREDICTED: Ovalbumin-like isoform X2 [Gavia stellata]	161	MVSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFHDKITGFEETIESQCSTSVSVHTSLKDMFT QITKPSDNYSLSFASRLYAEETYPILPEYLQCVKELYKGGLETISF QTAADQARELINSWVESQTDGMIKNILQPGSVDPQTEMVLVNA IYFKGMWEKAFKDEDTQAVPFRMTEQESKPVQMMYQIGSFKV AVMASEKMKILELPYASGGMSMLVMLPDDVSGLEQLETAITFE KLMEWTSSNMMEERKMKVYLPRMKMEEKYNLTSVLMALGM TDLFSSSANLSGISSAESLKMSEAVHEAFVEIYEAGSEAVGSTGA GMEVTSVSEEFRADHPFLFLIKHNPTNSILFFGRFCFSP
PREDICTED: Ovalbumin [Pelecanus crispus]	162	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFHDKITGFGEPIESQCGISVSVHTSLKDMITQ ITKPSDNYSLSFASRLYAEETYPILPEYLQCVKELYKGGLETISFQ TAADQARELINSWVENQTNNGMIKNILQPGSVDPQTEMVLVNAV YFKGMWEKAFKDEDTQAVPFRMTEQESKPVQMMYQIGSFKV AVMASEKMKILELPYASGELSMLVLLPDDVSGLEQLETAITLDK LTEWTSSNAMEERKMKVYLPRMKIEKKYNLTSVLIALGMTDLF SSSANLSGISSAESLKMSEAIHEAFLEIYEAGSEVVGSTEAGMEV TSVSEEFRADHPFLFLIKHNPTNSILFFGRCLSP
PREDICTED: Ovalbumin-like [Charadrius vociferus]	163	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFHDKITGSGTIEAQCGTSVSVHTSLKDMFT QITKPSDNYSVSFASRLYAEETYPILPEFLECVKELYKGGLESISF QTAADQARELINSWVESQTNNGMIKNILQPGSVDSQTEMVLVNA IYFKGMWEKAFKDEDTQTPFRMTEQETKPVQMMYQIGTFKV AVMPSEKMKILELPYASGELCMLVMLPDDVSGLEELESSITVEK LMEWTSSNMMEERKMKVFLPRMKIEEKYNLTSVLMALGMTD LFFSSANLSGISSAEPLKMSEAVHEAFIEIYEAGSEVVGSTGAGM EITSVSEEFRADHPFLFLIKHNPTNSILFFGRCLSP
PREDICTED: Ovalbumin-like [Eurypyga halias]	164	MGSIGAVSTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFHDKITGSGETIEAQCGTSVSVHTSLKDMFT QITKPSENYSVGFASRLYADETYPIIPEYLQCVKELYKGGLEMIS FQTAADQARELINSWVESQTNNGMIKNILQPGSVDPQTEMILVNA IYFKGVWEKAFKDEDTQAVPFRMTEQESKPVQMMYQFGSFKV AAMAAEKMKILELPYASGALSMLVLLPDDVSGLEQLESATIFE KLMEWTSSNMMEEEKKIKVYLPRMKMEEKYNFTSVLMALGMT DLFSSSANLSGISSADSLKMSEVVHEAFVEIYEAGSEVVGSTGS GMEAASVSEEFRADHPFLFLIKHNPTNSILFFGRFCFSP
PREDICTED: Ovalbumin-like isoform X1 [Gavia stellata]	165	MVSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFHDKITGFEETIESQVQKKQCSTSVSVHTSL KDMFTQITKPSDNYSLSFASRLYAEETYPILPEYLQCVKELYKG GLETISFQTAADQARELINSWVESQTDGMIKNILQPGSVDPQTE

		MVLVNAIYFKGMWEKAFKDEDTQAVPFRMTEQESKPVQMMYQIGSFKVAVMASEKMKILELPYASGGMSMLVMLPDDVSGLEQLETAITFEKLMWETSSNMMEERKMKVYLPRMKMEEKYNLTSVLMALGMTDLFSSSANLSGSSAESLKMSEAVHEAFVEIYEAGSEAVGSTGAGMEVTSVSEEFRADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin - like [<i>Egretta garzetta</i>]	166	MGSIGAASGEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLGARENTRAQIDKVVFHFDKIIGFGESIESQCGTSVSVHTSLKDMFAQITKPSDNYSLSFASRLYAEETFPILPEYLQCVKELYKGGLETLSFQTAADQARELINSWVESQTNGMIKDILQPGSVDPQTEMVLVNAIYFKGVWEKAFKDEDTQTVPFRMTEQESKPVQMMYQIGSFKVAVVAAEKIKILELPYASGALSMLVLLPDDVSSLEQLETAITFEKLTEWTSSNIMEERKIKVYLPRMKIEEKYNLTSVLMDLGITDLFSSSANLSGSSAESLKVSEAIHEAIVDIYEAGSEVVGSSGAGLEGTSVSEEFRADHPFLFLIKHNPTSSILFFGRCFSP
PREDICTED: Ovalbumin-like [<i>Balearica regulorum gibbericeps</i>]	167	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLGARENTRAQIDKVVFHFDKITGSGEAIESQCGTSVSVHISLKDMMFTQITKPSDNYSLSFASRLYAEETYPILPEYLQCVKELYKEGLATISFQTAADQAREFINSWVESQTNGMIKNILQPGSVDPQTMVVLVNAIYFKGVWEKAFKDEDTQAVPFRMTKQESKPVQMMYQIGSFKVAVMASEKMKILELPYASGQLSMLVMLPDDVSGLEQIENAITFEKLMWETNPNMMEERKMKVYLPRMKMEEKYNLTSVLMALGMTDLFSSSANLSGSSAESLKMSEAVHEAFVEIYEAGSEVVGSTGAGIEVTSVSEEFRADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [<i>Nestor notabilis</i>]	168	MGSIGEASTEFCIDVFRELKVQHVNNENIFYSPLSIISALSMVYLGARENTRAQIDQVVFHFDKITGFGDTVESQCGSSLSVHSSLKDIFAQITQPKDNYSLNFA SRLYAEETYPILPEYLQCVKELYKGGLETISFQTAADQARELINSWVESQTNGMIKNILQPSVDPQTEMVLVNAIYFKGVWEKAFKDEETQAVPFRITEQENRPVQIMYQFGSFKVAVVASEKIKILELPYASGQLSMLVLLPDEVSGLEQLENAITFEKLTEWTSSDIMEEKKIKVFLPRMKIEEKYNLTSVLVALGIADLFSSSANLSGSSAESLKMSEAVHEAFVEIYEAGSEVVGSSGAGIEAASDSEEFRADHPFLFLIKHKPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [<i>Pygoscelis adeliae</i>]	169	MGSIGAASTEFCFDIFNELKVQHVNNENIFYSPLSIISALSMVYLGARENKAQIDKVVFHFDKITGFGESIESQCSTASVHTSFKDMFTQITKPSDNYSLSFASRLYAEETYPILPEYSQC VKELYKGGLESISFQTAADQARELINSWVESQTNGMIKNILQPGSVDPQTELVLVNAIYFKGTWEKAFKDKDTQAVPFRVTEQESKPVQMMYQIGSYKVAVIASEKMKILELPYASGELSMLVLLPDDVSGLEQLETAITFEKLMWETSSNMMEERKVKVYLPRMKIEEKYNLTSVLMALGMTDLFSPSANLSGSSAESLKMSEAIHEAFVEIYEAGSEVVGSTEAGMEVTSVSEEFRADHPFLFLIKCNLTNSILFFGRCFSP
Ovalbumin-like [<i>Athene cunicularia</i>]	170	MGSISTASTEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLGARENTRAQIEKVVFHFDKITGFGESIESQCGTSVSVHTSLKDMLIQISKPSDNYSLSFASKLYAEETYPILPEYLQCVKELYKGGLESINFQTAADQARQLINSWVESQTNGMIKDILQPSVDPQTEMVLVNAIYFKGIWEKAFKDEDTQEV PFRITEQESKPVQMMYQIGSFKVAVIASEKIKILELPYASGELSMLIVLPDDVSGLEQLETAITFEKLIETSPSIMEERKTKVYLPRMKIEEKYNLTSVLMALGMTDLFSPSANLSGSSAESLKMSEAIHEAFVEIYEAGSEVVGSAEAGMEATSVSEFRVDHPFLFLIKHN PANIILFFGRCVSP

PREDICTED: Ovalbumin-like [Calidris pugnax]	171	MGSIGAASTEFCFDVFKELKVQHVNNENIFYSPLTIISALSLVYLG ARENTRAQIDKVFHFDKISGFGETTESQCGTSVSVHTSLKEMFT QITKPSDNYSVSFASRLYAEDTYPILPEYLQCVKELYKGGLETIS FQTAADQAREVINSWVESQTNGMIKNILQPGSVDSQTEMVLVN AIYFKGMWEKAFKDEDTQTMPFRITEQERKPVQMMYQAGSFK VAVMASEKMKILELPYASGEFCMLIMLPDDVSGLEQLENSFSFE KLMEWTTSNMMEERKMKVYIPRMKMEEKYNLTSVLMALGMT DLFSSSANLSGISSAETLKMSEAVHEAFMEIYEAGSEVVGSTGS GAEVTGVYEEFRADHPFLFLVKHKPTNSILFFGRCVSP
PREDICTED: Ovalbumin [Aptenodytes forsteri]	172	MGSIGAASTEFCFDIFNELKVQHVNNENIFYSPLSIISALSMVYLG ARENKAQIDKVVHFDKITGFGETIESQCSTSVSVHTSLKDTFTQ ITKPSDNYSLSFASRLYAEETYPILPEYSQCCKELYKGGLETISFQ TAADQARELINSWVESQTNGMIKNILQPGSVDPQTELVVNAIY FKGTWEKAFKDKDTQAVPFRVTEQESKPVQMMYQIGSYKVAV IASEKMKILELPYASRELSMLVLLPDDVSGLEQLETAITFEKLME WTSSNMMEERKVKVYLPRMKIEEKYNLTSVLMALGMTDLFSP SANLSGISSAESLKMSEAVHEAFVEIYEAGSEVVGSTGAGMEVT SVSEEFRADHPFLFLIKCNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Pterocles gutturalis]	173	MGSISAASAEFCLDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVHFDKITGSGETIEFQCGTSANIHPSLKDMFT QITRLSDNYSLSFASRLYAEERYPILPEYLQCVKELYKGGLETIS FQTAADQARELINSWVESQTNGMIKNILQPGSVNPQTEMVLVN AIYFKGLWEKAFKDEDTQTVPFMRMTEQESKPVQMMYQVGSFK VAVMASDKIKILELPYASGELSMLVLLPDDVTGLEQLETSITFEK LMEWTSSNVMEERTMKVYLPHMRMEEKYNLTSVLMALGVTD LFSSSANLSGISSAESLKMSEAVHEAFVEIYESGSQVVGSTGAGT EVTSVSEEFVRDHPFLFLIKHNPTNSILFFGRCFSP
Ovalbumin-like [Falco peregrinus]	174	MGSIGAASVEFCFDVFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENKAQIDKVVHFDKIAGFGAEIESQCVTASIIHSLKDMFTQI TKPSDNYSLSFASRLYAEAYPILPEYLQCVKELYKGGLETISFQ TAADQARDLINSWVESQTNGMIKNILQPGAVDLETEMVLVNAI YFKGMWEKAFKDEDTQTVPFMRMTEQESKPVQMMYQVGSFKV AVMASDKIKILELPYASGQLSMVVLLPDDVSGLEQLEASITSEK LMEWTSSSIMEEKKIKVYFPHMKIEEKYNLTSVLMALGMTDLF SSSANLSGISSAEKLVSEAVHEAFVEISEAGSEVVGSTEAGTEV TSVSEEFKADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin - like isoform X2 [Phalacrocorax carbo]	175	MGSIGAASSEFCFDIFKELKVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDKVVFPDKITASGESIESQCSTSVSVHTSLKDIFTQI TKSSDNHSLSFASRLYAEETYPILPEYLQCVKELYEGGLETISFQ TAADQARELINSWIESQTNGRIKNILQPGSVDPQTEMVLVNAIY FKGMWEKAFKDEDTQAVPFRMTEQESKPVQVMHQIGSFKVAV LASEKIKILELPYASGELSMLVLLPDDVSGLEQLETAITFEKLME WTSPNIMEERKIKVFLPRMKIEEKYNLTSVLMALGITDLFSP LANLSGISSAESLKMSEAIHEAFVEISEAGSEVIGSTEAEVEVTNDPE EFRADHPFLFLIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Merops nubicus]	176	MGSIGAASTEFCFDVFKELKAQYVNNENIFYSPMTIITALSMVYLG GSKENTRAQIAKVAHFDKITGFGESIESQCGASASIQFSLKDLFT QITKPSGNHSLSVASRIYAEETYPILPEYLECMKELYKGGLETIN FQTAANQARELINSWVERQTSGMIKNILQPSSVDSQTEMVLVN AIYFRGLWEKAFKVEDTQATPFRITEQESKPVQMMHQIGSFKV AVVASEKIKILELPYASGRLTMLVLLPDDVSGLKQLETTITFEKL

		MEWTTSNIMEERKIKVYLPRMKIEEKYNLTSVLMALGLTDLFSS SANLSGISSAESLKMSEAVHEAFVEIYEAGSEVVASAEAGMDAT SVSEEFRADHPFLFLIKDNTSNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Tauraco erythrolophus]	177	MGSIGAASTEFCDVFKELKGQHVNENIFFCPLSIVSALSMVYL GARENTRAQIVKVAHFDDKIAGFAESIESQCGTSVSIHTSLKDMFT QITKPSDNYSLNFAASRLYAEETYPPIPEYLQCVKELYKGGLETISF QTAADQAREIINSWVESQTNGMIKNILRPSSVHPQTELVLVNAV YFKGTWEKAFKDEDTQAVPFRITEQESKPVQMMYQIGSFKVA VTSEKMKILEVPYASGELSMLVLLPDDVSGLEQLETAITAELKIE WTSSTVMEERKLKVYLPRMKIEEKYNLTTVLTAALGVTDLFSSS ANLSGISSAQGLKMSNAVHEAFVEIYEAGSEVVGSKGEGTEVSS VSDEFKADHPFLFLIKHNPTNSIVFFGRCFSP
PREDICTED: Ovalbumin - like [Cuculus canorus]	178	MGSIGAASTEFCDVFKELKVHHVNENILYSPLAISALSMVYLG AKENTRDQIDKVHFDKITGIGESIESQCSTAVSVHTSLKDVFD QITRPSDNYSLAFASRLYAEKTYPIPEYLQCVKELYKGGLETID FQTAADQARQLINSWVEDETNGMIKNILRPSSVNPQTKILVNAI YFKGMWEKAFKDEDTQEVFPFRITEQETKSVQMMYQIGSFKVA EVVSDKMKILELPYASGKLSMLVLLPDDVYGLEQLETVITVEKL KEWTSSIVMEERITK VYLPRMKIMEKYNLTSVLTAFGITDLFSPS ANLSGISSTESLKVSEAVHEAFVEIHEAGSEVVGSAAGAGIEATSV SEEFKADHPFLFLIKHNPTNSILFFGRCFSP
Ovalbumin [Antrostomus carolinensis]	179	MGSIGAASTEFCLDVFKELKVQHVNENIFYSPLSIISALSMVYLG ARENTRAQIDKVHFDKITGFEDSIESQCGTSVSVHTSLKDMFT QITKPSDNYSVGFASRLYAAETYQILPEYSQCVKELYKGGLETI NFQKAADQATELINSWVESQTNGMIKNILQPSSVDPQTQIFLVN AIYFKGMWQRAFKEEDTQAVPFRISEKESKPVQMMYQIGSFKV AVIPSEKIKILELPYASGLLSMLVILPDDVSGLEQLENAITLKLKLM QWTSSNMMEERKIKVYLPRMRMEEKYNLTSVFMALGITDLFSS SANLSGISSAESLKMSDAVHEASVEIHEAGSEVVGSTGSGTEAS SVSEEFRADHPYLFLIKHNPTDSIVFFGRCFSP
PREDICTED: Ovalbumin-like [Opisthocomus hoazin]	180	MGSIGAASTEFCDVFKELKFQHVNDENIFYSPLSIISALSMVYLG ARENTRAQIDKVHFDKIAGFEETVESQCGTSVSVHTSLKDMF AQITKPSDNYSLSFASRLYAEETYPILPEYLQCVKELYKGGLETI SFQTAADQARDLINSWVESQTNGMIKNILQPSSVGPQTELILVN AIYFKGMWQKAFKDEDTQEVPFMRMTEQQSKPVQMMYQTGSF KVAVVASEKMKILALPYASGQLSLLVMLPDDVSGLKQLESAT SEKLIEWTSPSMMEERKIKVYLPRMKIEEKYNLTSVLMALGITD LFSPSANLSGISSAESLKMSSQAVHEAFVEIYEAGSEVVGSTGAG MEDSSDSEEFVVDHPFLFFIKHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Lepidothrix coronata]	181	MGSIGPLSVEFCCDVFKELRIQHPRNIFYSPVTIISALSMVYLG RDNTKAQIEKAVHFDKIPGFESIESQCGTSLSIHTSLKDIFTQIT KPSDNYSYTVGIASRLYAEKYPILPEYLQCIKELYKGGLEPINFQT AAEQARELINSWVESQTNGMIKNILQPSSVNPETDMVLVNAIYF KGLWEKAFKDEDIQTVPFRITEQESKPVQMMMFQIGSFRVAEITS EKIRILELPYASGQLSLWVLLPDDISGLEQLETAITFENLKEWTSS TKMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANLSGI SSAESLKVSSAFHEASVEIYEAGSKVVGSTGAEVEDTSVSEEF ADHPFLFLIKHNPSNSIFFFGRCFSP
PREDICTED: Ovalbumin [Struthio]	182	MGSIGTASAEFCFDVFKELKVHHVNENIFYSPLSIISALSMVYLG ARENTKTQMEKVIHFDKITGLGESMESQCGTGVSIIHTALKDML SEITKPSDNYSLSLASRLYAEQTYAILPEYLQCIKELYKESLETVS

camelus australis]		FQTAADQARELINSWIESQTNGVIKNFLQPGSVDSQTELVVNA IYFKGMWEKAFKDEDTQEVPRITEQESRPVQMMYQAGSFKV ATVAAEKIKILELPYASGELSMLVLLPDDISGLEQLETTISFEKLT EWTSSNMEDRNMKVYLPRMKIEEKYNLTSVLIALGMTDLFSP AANLSGISAAESLKMSEAIHAAAYVEIYEADSEIVSSAGVQVEVT SDSEEFRVDHPFLFLIKHNPTNSVLFFGRCISP
PREDICTED: Ovalbumin-like [Acanthisitta chloris]	183	MGSIGAVSTEFSCDVFKELRIHHVQENIFYSPVTIISALSMIYLGA RDSTKAQIEKAVHFDKIPGFGESESQCSTLSIHTSIKDMFTKIT KASDNYSIGIASRLYAEKYPILPEYLQCVKELYKGGLESISFQT AAEQAREIINSWVESQTNGMIKNILQPSSVDPQTDIVLVNAIYFK GLWEKAFRDEDTQTVPFKITEQESKPVQMMYQIGSFKVAEITSE KIKILEVPYASGQLSLWVLLPDDISGLEKLETAITFENLKEWTSS TKMEERKIKVYLPRMKIEEKYNLTSVLTALGITDLFSSSANLSGI SSAESLKVSEAFHEAIVEISEAGSKVVGSGAGVDDTSVSEEFR ADHPFLFLIKHNPTSSIFFFGRCFSP
PREDICTED: Ovalbumin-like [Tyto alba]	184	MGSIGAASTEFCDVFKELKVQHVNNENIFYSPLSIISALSMVYLGA ARENTRAQIDKVVFHDKIAGFGESTESQCSTSVSAHTSLKDMS NQITKLSDNYSLSFASRLYAEETYPILPEYSQCVELYKGGLESIS SFQTAAYQARELINAWVESQTNGMIKDILQPGSVDSQTKMVLV NAIYFKGIWEKAFKDEDTQEVPRMTEQETKPVQMMYQIGSFK VAVIAAEKIKILELPYASGQLSMLVILPDDVSGLEQLETAITFEK LTEWTSASVMEERKIKVYLPRMSIEEKYNLTSVLIALGVTDLFS SSANLSGISSAESLRMSEAIHEAFVETYEAGSTESGTEVTSASEE FRVDHPFLFLIKHKPTNSILFFGRCFSP
PREDICTED: Ovalbumin - like isoform X1 [Phalacrocorax carbo]	185	MGSIGAAASSEFCDFIKELKVQHVNNENIFYSPLSIISALSMVYLGA ARENTRAQIDKVVPFDKITASGESIESQVQKIQCSTSVSVHTSLK DIFTQITKSSDNHSLSFASRLYAEETYPILPEYLQCVKELYEGGL ETISFQTAADQARELINSWIESQTNGRIKNILQPGSVDPQTEMVL VNAIYFKGMWEKAFKDEDTQAVPFRMTEQESKPVQVMHQIGS FKVAVLASEKIKILELPYASGELSMLVLLPDDVSGLEQLETAITF EKLMEWTSPNIMEERKIKVFLPRMKIEEKYNLTSVLMALGITDL FSPLANLSGISSAESLKMSEAIHEAFVEISEAGSEVIGSTEAEVEV TNDPEEFRADHPFLFLIKHNPTNSILFFGRCFSP
Ovalbumin-like [Pipra filicauda]	186	MGSIGPLSVEFCCDVFKELRIQHARENIFYSPVTIISALSMVYLGA ARDNTKAQIEKAVHFDKIPGFGESESQCSTLSIHTSLKDIFTQI TKPSDNYTVGIASRLYAEKYPILPEYLQCIKELYKGGLEPISFQ TAAEQARELINSWVESQTNGIKNILQPSSVNPETDMVLVNAIYF KGLWEKAFKDEGTQTVPFRITEQESKPVQMMFQIGSFRVAEIAS EKIRILELPYASGQLSLWVLLPDDISGLEQLETAITFENLKEWTSS TKMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANLSGI SSAERLKVSSAFHEASMEINEAGSKVVGAGVDDTSVSEEFRVD RPFLFLIKHNPSNSIFFFGRCFSP
Ovalbumin [Dromaius novaehollandiae]	187	MGSIGAASTEFCDMFKELKVHHVNNENIISPLSIISILSMVFLGA RENTKTQMEKVIHFDKITGFGESESQCSTSVSVHASLKDILSEI TKPSDNYSLSLASKLYAEETYPVLPEYLQCIKELYKGSLETVSFQ TAADQARELINSWVETQTNGVIKNFLQPGSVDPQTEMVLVDAI YFKGTWEKAFKDEDTQEVPRITEQESKPVQMMYQAGSFKVA TVAAEKMKILELPYASGELSMFVLLPDDISGLEQLETTISIEKLSE WTSSNMEDRKMKVYLPHMKIEEKYNLTSVLVALGMTDLFSP SANLSGISTAQTLKMSEAIHGAYVEIYEAGSEMATSTGVLVEAA SVSEEFRVDHPFLFLIKHNPSNSILFFGRCIFP

Chain A, Ovalbumin	188	MGSIGAASTEFCFDMFKELKVHHVNNENIYSPLSIISILSMVFLGAR RENTKTQMEKVIHFDKITGFGESLESQCGTSVSVHASLKDILSEI TKPSDNYSLSLASKLYAEETYPVLPEYLQCIKELYKGSLETVSFQ TAADQARELINSWVETQTNGVIKNFLQPGSVPQTEMVLVDAI YFKGTWEKAFKDEDTQEVPRITEQESKPVQMMYQAGSFKVA TVAAEKMKILELPYASGELSMFVLLPDDISGLEQLETTISIEKLSE WTSSNMEDRKMKVYLPHMKIEEKYNLTSVLVALGMTDLFSP SANLSGISTAQTLKMSEAIHGAYVEIYEAGSEMATSTGVLVEAA SVSEEFVRVDHPFLFLIKHNPSNSILFFGRCIFPHHHHH
Ovalbumin-like [Corapipo altera]	189	MGSIGPLSVEFCCDVFKELRIQHARENIFYSPVTIISALSMVYLG ARDNTKAQIEKAVHFDKIPGFGESIESQCGTSLSIHTSLKDIFTQI TKPSDNYTVGIASRLYAEKYPILPEYLQCIKELYKGGLEPISFQ TAAEQARELINSWVESQTNGMIKNILQPSAVNPETDMVLVNAI YFKGLWEKAFKDEGTQTPFRITEQESKPVQMMFQIGSFRVAEI TSEKIRILELPYASGQLSLWLLPDDISGLEQLETAITFENLKEW TSSTKMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANL SGISSAERLKVSSAFHEASMEIYEAGSKVVGSTGAGVDDTSVSE EFRVDRPFLFLIKHNPSNSIFFFGRCFSP
Ovalbumin-like protein [Amazona aestiva]	190	MEDQRGNTGFTMGSIGAASTEFCIDVFREL RVQHVNNENIFYSPL TIISALSMVYLGARENTRAQIDQVVHFDKIAGFGDTVESQCGSS PSVHNSLKT VXAQITQPRDNYSLNLASRLYAEESYPILPEYLQC VKELYNGGLETVSFQTAADQARELINSWVESQTNGIKNILQPS VDPQTEMVLVNAIYFKGLWEKAFKDEETQAVPFRITEQENRPV QMMYQFGSFKVAXVASEKIKILELPYASGQLSMLVLLPDEVSG LEQNAITFEKLTEWTSSDLMEERKIKVFFPRVKIEEKYNLTAVL VSLGITDLFSSSANLSGISSAENLKMSEAVHEAXVEIYEAGSEVA GSSGAGIEVASDSEEFVRVDHPFLFLIXHNPTNSILFFGRCFSP
PREDICTED: Ovalbumin-like [Melopsittacus undulatus]	191	MGSIGAASTEFCIDVFREL RVQHVNNENIFYSPLSIISALSMVYLG ARENTRAQIDEVFHFDKIAGFGDTVDPQCGASLSVHKS LQNVF AQITQPKDNYSLNLASRLYAEESYPILPEYLQCVKELYNEGLET VSFQTGADQARELINSWVENQTNGVIKNILQPSVDPQTEMVL VNAIYFKGLWQKAFKDEETQAVPFRITEQENRPVQMMYQFGSF KVAVVASEKVKILELPYASGQLSMWVLLPDEVSGLEQLENAIT FEKLTEWTSSDLTEERKIKVFLPRVKIEEKYNLTAVLMALGVTD LFSSSANFSGISAAENLKMSEAVHEAFVEIYEAGSEVVGSSGAGI EAPSDSEEFRADHPFLFLIKHNPTNSILFFGRCFSP
Ovalbumin-like [Neopelma chrysocephalum]	192	MGSIGPLSVEFCCDVFKELRIQHARDNIFYSPVTIISALSMVYLG ARDNTKAQIEKAVHFDKIPGFGESIESQCGTSLSVHTSLKDIFTQI TKPRENYTVGIASRLYAEKYPILPEYLQCIKELYKGGLEPISFQ TAAEQARELINSWVESQTNGMIKNILQPSVNPETDMVLVNAIY FKGLWKKAFKDEGTQTPFRITEQESKPVQMMFQIGSFRVAEIT SEKIRILELPYASGQLSLWLLPDDISGLEQLES AITFENLKEWTS STKMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANLSG ISSAEKLVSSAFHEASMEIYEAGNKVVGSTGAGVDDTSVSEEF RVDRPFLFLIKHNPSNSIFFFGRCFSP
PREDICTED: Ovalbumin-like [Buceros rhinoceros silvestris]	193	MGSIGAAAEFCVDVFKEKLDQHVNNIVFSPLMIISALSMVNIG AREDTRAQIDKVHFDKITGYGESIESQCGTSIGIYFSLKDAFTQI TKPSDNYSLSFASKLYAEETYPILPEYLKCVKELYKGGLETISFQ TAADQARELINSWVESQTNGMIKNILQPSVDPQTEMVLVNAIY FKGLWEKAFKDEDTQAVPFRITEQESKPVQMMYQIGSFKVAVI ASEKIKILELPYASGQLSLLVLLPDDVSGLEQLES AITSEKLLW

		TNPNIMEERKTKVYLPRMKIEEKYNLTSVLVALGITDLFSSSAN LSGISSAEGKLSDAVHEAFVEIYEAGREVVGSSSEAGVEDSSVS EEFKADRPFIKHNPTNGILYFGRYISP
PREDICTED: Ovalbumin-like [Cariama cristata]	194	MGSIGAANTDFCFDVFKEKLVHHANENIFYSPLSIVSALAMVYL GARENTRAQIDKALHFDKILGFGETVESQCDTSVSVHTSLKDM LIQITKPSDNYSFSFASKIYTEETYPILPEYLQCVKELYKGGVETI SFQTAADQAREVINSWVESHTNGMIKNILQPGSVDPQTKMVLV NAVYFKGIWEKAFKEEDTQEMPFRINEQESKPVQMMYQIGSFK LTVAASENLKILEFPYASGQLSMMVILPDEVSGLKQLET SITSEK LIKWTSSNTMEERKIRVYLPRMKIEEKYNLKSVMALGITDLFS SSANLSGISSAESLKMSEAVHEAFVEIYEAGSEVTSSTGTEMEAE NVSEEFKADHPFLFLIKHNPTDSIVFFGRCMSP
Ovalbumin [Manacus vitellinus]	195	MGSIGPLSVEFCCDVFKEKLRHQHARENIFYSPVTIISALSMVYLGA ARDNTKAQIEKAVHFDKIPGFGESIESQCGTSLSIHTSLKDIFTQI TKPSDNYTVGIASRLYAEKYPILPEYLQCIKELYKGGLEPISFQ TAAEQARELINSWVESQTNGMIKNILQPSSVNPETDMVLVNAIY FKGLWEKAFKDESTQTVPFRITEQESKPVQMMFQIGSFRVAEIA SEKIRILELPYASGQLSLWVLLPDDISGLEQLETAITFENLKEWTS STKMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANLSG ISSAERLKVSSAFHEASMEIYEAGSRVVEAGVDDTSVSEEFVRD RPFLFLIKHNPSNSIFFFFGRCFSP
Ovalbumin-like [Empidonax traillii]	196	MGSIGPVSTEFCCDIFKEKLRHQHARENIYSPVTIISALSMVYLGA RDNTKAQIEKAVHFDKIPGFGESIESQCGTSLSIHTSLKDILTQIT KPSDNYTVGIASRLYAEKYPILSEYLQCIKELYKGGLEPISFQT AAEQARELINSWVESQTNGMIKNILQPSSVNPETDMVLVNAIYF KGLWEKAFKDEGTQTVPFRITEQESKPVQMMFQIGSFKVAEITS EKIRILELPYASGKLSLWVLLPDDISGLEQLETAITFENLKEWTSS TRMEERKIKVYLPRMKIEEKYNLTSVLTSLGITDLFSSSANLSGI SSAERLKVSSAFHEVFVEIYEAGSKVEGSTGAGVDDTSVSEEFR ADHPFLFLVKHNPSNSIFFFFGRCYLP
PREDICTED: Ovalbumin-like [Leptosomus discolor]	197	MGSTGAASMEFCFALFRELKVQHVNENIFFSPVTIISALSMVYL GARENTRAQLDKVAPFDKITGFGETIGSQCSTSASSHTSLKDV TQITKASDNYSLSFASRLYAEETYPILPEYLQCVKELYKGGLES SFQTAADQARELINSWVESQTNGMIKDILRPSSVDPQTKIILITAI YFKGMWEKAFKEEDTQAVPFRMTEQESKPVQMMYQIGSFKVA VIPSEKLIKILELPYASGQLSMLVILPDDVSGLEQLETAITTEKLKE WTSPSMMKERKMKVYFPRMRIEEKYNLTSVLMA LGITDLFSPS ANLSGISSAESLKVSEAVHEASVDIDEAGSEVIGSTGVGTEVTSV SEEIRADHPFLFLIKHKPTNSILFFGRCFSP
Hypothetical protein H355_008077 [Colinus virginianus]	198	MEHAQLTQLVNSNMTSNTCHEADEFENIDFRMDSISVTNTKFC FDVFNEMKVHHVNENILYSPLSILTALAMVYLGARGNTESQMK KALHFDSITGAGSTTDSQCGSSEYIHNLFKEFLTEITRTNATYSL EIADKLYVDKTFTVLPEYINCARKFYTGGEVEEVNFKTAAEEAR QLINSWVEKETNGQIKDLLVPSSVDFGTMMVFINTIYFKGIWKT AFNTEDTREMPPFSMTKQESKPVQMMCLNDTFNMATLPAEKMR ILELPYASGELSMLVLLPDEVSGLEQIEKAINFEKLREWTSNA MEKKSMKVYLPRMKIEEKYNLTSTLMALGMTDLFSSSANLTGI SSVENLMISDAVHGAFMEVNEEGTEAAGSTGAIGNIKHSVEFEE FRADHPFLFLIRYNPTNVILFFDNSEFTMG SIGAVSTEFCDVFK ELRVHHANENIFYSPFTVISALAMVYLGAKDSTRTQINKVVRFD KLPGFGDSIEAQCGTSANVHSSLRDILNQITKPNDIYSFSLASRL

		YADETYTILPEYLQCVKELYRGGLESINFQTAADQARELINSWV ESQTSGIIRNVLPSSVDSQTAMVLVNAIYFKGLWEKGFKDED QAMPFRVTEQENKSVQMMYQIGTFKVASVASEKMKILELPFAS GTMSMWVLLPDEVSGLEQLETTISIEKLTEWTSSSVMEERKIKV FLPRMKMEEKYNLTSVLMAMGMTDLFSSSANLSGISSTLQKKG FRSQELGDKYAKPMLESPALTPQVTAWDNSWIVAHPAAIEPDL CYQIMEQKWKPFDWPDFRLPMRVSCRFRTEALNKANTSFAL DFFKHECQEDDDENILFSPFSSSALATVYLGAKGNTADQMAKT EIGKSGNIHAGFKALDLEINQPTKNYLLNSVNQLYGEKSLPFSK EYLQLAKKYSAEPQSVDFLGKANEIRREINSRVEHQTEGKIKN LLPPGSIDSLTRLVLVNALYFKGNWATKFEAEDTRHRPFRINMH TTKQVPMMYLRDKFNWYVESVQTDVLELPYVNNDLSMFILL PRDITGLQKLINELTFEKLAWTSPELMEKMKMEVYLPRFTVEK KYDMKSTLSKMGIEDAFTKVDSCGVTNVDEITTHIVSSKCLELK HIQINKKLKCNKAVAMEQVSASIGNFTIDLFNKLNETSRDKNIFF SPWSVSSALALTSLAAKGNTAREMAEDPENEQAENIHSGFKEL MTALNKPRNTYSLKSANRIYVEKNYPLLPTYIQLSKKYKAEP YKVNFKTAPEQSRKEINNWEKQTERKIKNFLSSDDVKNSTKSI LVNAIYFKAEEWEEKFQAGNTDMQPFRMSKNKSKLVKMMYMR HTFPVLIMEKLNFKMIELPYVKRELSMFILLPDDIKDSTTGLEQL ERELTYEKLSEWADSKKMSVTLVDLHLPKFSMEDRYDLKDAL KSMGMASAFNSNADFGMTGFGQAVPMESLSASTNSFTLDLYK KLDETSKGQNIFFASWSIATALAMVHLGAKGDTATQVAKGPEY EETENIHSGFKELLSAINKPRNTYLMKSANRLFGDKTYPLLPKFL ELVARYYYQAKPQAVNFKTDAEQARAQINSWVENETESKIQNLL PAGSIDSHTVLVLVNAIYFKGNWEKRFLEKDTSKMPFRLSKTET KPVQMMFLKDTFLIHHERTMKFKIIELPYVGNELSAFVLLPDDIS DNTTGLELVERELTYEKLAEWSNSASMMKAKVELYLPKLKME ENYDLKSVLSDMGIRSAFDPAQADFTRMSEKKDLFISKVIHKA FEVNEEDRIVQLASGRLTGRCTRLANKELSEKNRTKNLFFSPFSI SSALSMILLGSKGNTEAQIAKVLSSKAEDAHNQYQSLLSEINN PDTKYILRTANRLYGEKTFEFLSSFIDSSQKFYHAGLEQTDFKNA SEDSRKQINGWVEEKTEGKIQKLLSEGIINSMTKLVLVNAIYFK GNWQEKFDKETTKEMPFKINKNETKPVQMMFRKGKYNMTYIG DLETTVLEIPYVDNELSMILLPDSIQDESTGLEKLERELTYEKL MDWINPNMMDSTEVRVSLPRFKLEENYELKPTLSTMGMPDAF DLRTADFGSGISSGNELVLSEVVHKSFEVNEEGTEAAAATAGIM LLRCAMIVANFTADHPFLFFIRHNKTNLSILFCGRFCSP
PREDICTED: Ovalbumin isoform X2 [Apteryx australis mantelli]	199	MGSIGTASTEFCDFMFKEMKVQHANQNIIFSPLTIISALSMVYLG ARDNTKAQMEKVIHFDKITGFGESVESQCGTSVSIHTSLKDMLS EITKPSDNYSLSLASRLYAEETYPILPEYLQCMKELYKGGLETVS FQTAADQARELINSWVESQTNGVIKNFLQPGSVDPQTEMLVN AIYFKGMWEKAFKDEDTEQVPFRITEQESKPVQMMYQVGSFK VATVAAEKMKILEIPYTHRELSMFVLLPDDISGLEQLETTISFEK LTEWTSSNMMEERKVKVYLPHMKIEEKYNLTSVLMALGMTDL FSPSANLSGISTAQTLMMSEAIHGAYVEIYEAGREMASSTGVQV EVTSVLEEVRADKPFLFFIRHNPTNSMVVFGRYMSP
Hypothetical protein ASZ78_006007	200	MTSNTCHEADEFENIDFRMDSISVTNTKFCFDVFNEMKVHHVN ENILYSPLSILTALAMVYLGARGNTESQMKKALHFDSITGGGST TDSQCGSSEYIHNLFKEFLTEITRTNATYSLEIADKLYVDKFTTV LPEYINCARKFYTGGVVEEVNFKTAAEEARQLMNSWVEKETNG

[Callipepla squamata]		QIKDLLVPSSVDFGTMMVFINTIYFKGIWKTAFTNTEDTREMPPFS MTKQESKPVQMMCLNDTFNMVTLPAEKMRIELPYASGELSM LVLLPDEVSGLERIEKAINFEKLREWSTSTNAMEKKSMKVYLPR MKIEEKYNLTSTLMALGMTDLFSRSANLTGISSVDNLMISDAVH GAFMEVNEEGTEAAGSTGAIGNIKHSVEFEEFRADHPFLFLIRY NPTNVILFFDNSEFTMG SIGAVSTEFCDVFKELRVHHANENIFY SPFTIISALAMVYLGAKDSTRTQINKVVRFDKLPFGGDSIEAQCG TSANVHSSLRDILNQITKPNDIYSFSLASRLYADETYTILPEYLQC VKELYRGGLESINFQTAADQARELINSWVESQTSIGIRNVLPQSS VDSQTAMVLVNAIYFKGLWEKGFKDEDTQAIPFRVTEQENKSV QMMYQIGTFKVASVASEKMKILELPFASGTMSMWVLLPDEV GLEQLETTISIEKLTEWTSSSVMEERKIKVFLPRMKMEEKYNLT SVLMAMGMTDLFSSSANLSGISSTLQKKGFRSQELGDKYAKPM LESPALTPQATAWDNSWIVAHPPAIEPDLYYQIMEQKWKPFDW PDFRLPMRVSCRFRTEALNKANTSFAIDFFKHECQEDDSENIL FSPFSSSALATVYLGAKGNTADQMAKVLHFNEAEGARNVTTTI RMQVYSRTDQQRLNRRACFQKTEIGKSGNIHAGFKGLNLEINQ PTKNYLLNSVNQLYGEKSLPFSKEYLQLAKKYSAEPQSVDFV GTANEIRREINSRVEHQTEGKIKNLLPPGSIDSLTRLVLVNALYF KGNWATKFEAEDTRHRPFRINTHTTKQVPMMYLSDKFNWTVY ESVQTDVLELPYVNNDLSMFILLPRDITGLQKLINELTFEKL SAWTSPELMEKMKMEVYLPRFTVEKKYDMKSTLSKMGIEDAFTKV DNCGVTNVDEITIHVPSKCLELKHQINKELKCNKAVAMEQVS ASIGNFTIDLFNKLNETSRDKNIFFSPWSVSSALALTS LAAGNTAREMAEDPENEQAENIHSGFNELLTALNKPRNTYSLK SANRIYVEKNYPLLPTYIQLSKKYKAEPHKVNFKTAPEQSRKEIN NWVEKQTERKIKNFLSSDDVKNSTKLILVNAIYFKA EWEKFKQAGNTDMQPFRMSKNKSKLVKMMYMRHTFPVL IMEKLNFKMIELPYVKRELSMFILLPDDIKDSTTGLEQLEREL TYEKLSEWADSKKMSVTLVDLHLPKFSMEDRYDLKD ALRSMGMASAFNSNADFSGMTGERDLVISKVCHQSFV AVDEKGTEAAAATAVIAEAVPMESLSASTNSFTLDLY KKLDETSKGQNIFFASWSIATALTMVHLGAKGDTA TQVAKGPEYEETENIHSGFKELLSALNKPRNTYS MKSANRLFGDKTYPLLPTKTKPVQMMFLKDTFLI HHERTMKFKIIELPYMGNELSAFVLLPDDISDNTT GLELVERELTYEKLAEWSNSASMMKVVELYLPKL KMEENYDLKSALSDMGIRSAFDPAQADFTRMSEKK DLFISKVIHKAFVEVNEEDRIVQLASGRLTGNT EAQIAKVLSSK AEDAHNGYQSLLEINNPDTKYIL RTANRLYGEKTFEFLSSFIDSQKFYHAGLEQTDF KNASEDSRKQINGWVEEKTEGKIQLLSEGIINS MTKLVLVNAIYFKGNWQEKFDKETTKEMP FKINKNETKPVQMMFRKGKYNMTYIGDLETT VLEIPYVDNELSMIILLPDSIQDESTGLEKLEREL TYEKLMDWINPNMMDSTEVRVSLPRFKLEEN YELKPTLSTMGMPDAFDLRTADFSGISSG NELVLSEVVHKSFVEVNEEGTEAAAATAGIM LLRCAMIVANFTADHPFLFFIRHNKTNSILFC GRFCSP
PREDICTED: Ovalbumin-like [Mesitornis unicolor]	201	MASIGAASTEFCFDVFKELKTQHVKENIFYSPMAIISALSMVYIG ARENTRAIDKVVFHDKITGFGNAVESQCGPSVSVHSSLKDLIT QISKRSNYSLSYASRIYAEETYPILPEYLQCVKEVYKGGLESISF QTAADQARENINAWVESQTNGMIKNILQSSVNPQTEMVLVNA IYLKGMWEKAFKDEDTQTMPFRVTQQESKPVQMMYQIGSFKV AVIASEKMKILELPYTSGQLSMLVLLPDDVSGLEQVESAITAEK

		LMEWTPSPSIMEERTMKVYLPRMKMVEKYNLTSVLMALGMTD LFTSVANLSGISSAQGLKMSQAIHEAFVEIYEAGSEAVGSTGVG MEITSVSEEFKADLSFLFLIRHNPTNSIIFFGRCISP
Ovalbumin, partial [Anas platyrhynchos]	202	MGSIGAASTEFCDVRELRVQHVNNENIFYSPFSIISALAMVYLG ARDNTRTQIDKISQFQALSDEHLVLCIQQLGEFFVCTNRREV TRYSEQTEDKTQDQNTGQIHKIVDTCMLRQDILTQITKPSDNFS LSFASRLYAEETYAILPEYLQCVKELYKGGLESISFQTAADQAR ELINSWVESQTNGIKNILQPSSVDSQTTMVLVNAIYFKGMWEK AFKDEDTQAMPFRMTEQESKPVQMMYQVGSFKVAMVTSEKM KILELPFASGMMSMFVLLPDEVSGLEQLESTISFEKLTEWTSSTM MEERRMKVYLPRMKMEEKYNLTSVFMALGMTDLFSSSANMS GISSTVSLKMSEAVHAACVEIFEAGRVDVGSAGMDVTSVSE EFRADHPFLFFIKHNPTNSILFFGRWMSP
PREDICTED: Ovalbumin-like [Chaetura pelagica]	203	MGSIGAAAEFCLDIFKELKVQHVNNENIIFSPMTIISALSLVYLGA KEDTRAQIEKVVPFDKIPGFGEIVESQCPKSASVHSSIQDIFNQIK RSDNYSLSLASRLYAEESYPIRPEYLQCVKELDKEGLETISFQTA ADQARQLINSWVESQTNGMIKNILQPSSVNSQTEMVLVNAIYFR GLWQKAFKDEDTQAVPFRITEQESKPVQMMQIGSFKVAEIAS EKMKILELPYASGQLSMLVLLPDDVSGLEKLESSITVEKLIWTS SNLTEERNVKVYLPRLKIEEKYNLTSVLAALGITDLFSSSANLSG ISTAESLKLRAVHESFVEIQEAGHEVEGPKEAGIEVTSALDEFR VDRPFLFVTKHNPTNSILFLGRCLSP
PREDICTED: Ovalbumin-like [Apaloderma vittatum]	204	MGSISAASGEFCLDIFKELKVQHVNNENIFYSPMVIVSALSLSVYLGA ARENTRAQIDKVIPFDKITGSSEAVESQCGTPVGAHISLKDVFQA IAKRSDNYSLSFVNRLYAEETYPIPEYLQCVKELYKGGLETISF QTAADQAREIINSWVESQTDGKIKNILQPSSVDPQTKMVLVSAI YFKGLWEKSFKDEDTQAVPFRVTEQESKPVQMMYQIGSFKVA AIAAEKIKILELPYASEQLSMLVLLPDDVSGLEQLEKKISYEKLT EWTSSSVMEKKIKVYLPRMKIEEKYNLTSILMSLGITDLFSSSA NLSGISSTKSLKMSEAVHEASVEIYEAGSEASGITGDGMEATSV FGEFKVDHPFLFMKHKPTNSILFFGRCISP
Ovalbumin-like [Corvus cornix cornix]	205	MGSIGPVSTEVCCDIFRELRSQSVQENVCYSPLLIISTLSMVYIGA KDNTKAQIEKAHFDPKIPGFGESTESQCGTSVSIHTSLKDIFTQIT KPSDNYSISIIARRLYAEKYPILPEYIQCVKELYKGGLESISFQTA AEKSRELINSWVESQTNGTIKNILQPSSVSSQTDMLVLSAIYFKG LWEKAFKEEDTQTIPFRITEQESKPVQMMSQIGTFKVAEIPSEKC RILELPYASGRLSLWLLPDDISGLEQLETAITFENLKEWTSSSK MEERKIRVYLPRMKIEEKYNLTSVLKSLGITDLFSSSANLSGISS AESLKVSAAFHEASVEIYEAGSKGVGSSEAGVDGTSVSEEIRAD HPFLFLIKHNPSDSILFFGRCFSP
PREDICTED: Ovalbumin-like [Calypte anna]	206	MGSIGAASTEFCDVFKELKVQHVNNENIISPLSIISALSMVYIGA REDTRAQIDKVVHFDKITGFGEAIESQCPTSESVHASLKETFSQL TKPSDNYSLSLAFASRLYAEETYPIPEYLQCVKELYKGGLETINFQ TAAEQARQVINSWVESQTDGMIKSLQPSVDPQTEMILVNAIY FRGLWERAFAKDEDTQELPFRITEQESKPVQMMSQIGSFKVAVV ASEKVKILELPYASGQLSMLVLLPDDVSGLEQLESSITVEKLIW ISSNTKEERNIKVYLPRMKIEEKYNLTSVLVALGITDLFSSSANL SGISSAESLKISEAVHEAFVEIQEAGSEVVGSPGPEVEVTSVSEE WKADRPFLFLIKHNPTNSILFFGRYISP
PREDICTED: Ovalbumin	207	MGSIGPVSTEVCCDIFRELRSQSVQENVCYSPLLIISTLSMVYIGA KDNTKAQIEKAHFDPKIPGFGESTESQCGTSVSIHTSLKDIFTQIT

[Corvus brachyrhynchos]		KPSDNYSISIIARRLYAEEKYPILQEYIQCVKELYKGGLESISFQTA AEKSRELINSWVESQTNGTIKNILQPSSVSSQTDMLVLSAIYFKG LWEKAFKEEDTQTIPFRITEQESKPQVQMMSQIGTFKVAEIPSEKC RILELPYASGRLSLWVLLPDDISGLEQLETSITFENLKEWTSSSK MEERKIRVYLPRMKIEEKYNLTSVLKSLGITDLFSSSANLSGISS AESLKVSAVFHEASVEIYEAGSKGVGSSEAGVDGTSVSEEIRAD HPFLFLIKHNPSDSILFFGRCFSP
Hypothetical protein DUI87_08270 [Hirundo rustica rustica]	208	MLNLMHPKQFCCTMGSIGPVSTEVCCDIFRELRSQSVQENVCY SPLLIISTLSMVYIGAKDNTKAQIEKAHFDPKIPGFGESESTESQCGT SVSIHTSLKDIFTQITKPSDNYSISIASRLYAEEKYPILPEYIQCVK ELYKGGLESISFQTA AEKSRELINSWVESQTNGTIKNILQPSSVSS QTDMLVLSAIYFKGLWEKAFKEEDTQTPFRITEQESKPQVQMM SQIGTFKVAEIPSEKCRILELPYASGRLSLWVLLPDDISGLEQLET AITSENLKEWTSSSKMEERKIKVYLPRMKIEEKYNLTSVLKSLGI TDLFSSSANLSGISSAESLKVSGAFHEAFVEIYEAGSKAVGSSGA GVEDTSVSEEIRADHPFLFFIKHNPSDSILFFGRCFSP
Ostrich OVA sequence as secreted from pichia	209	EAEAGSIGTASAEFCFDVFKELKVHHVNENIFYSPLSIISALSMV YLGARENKTQMEKVIHFDKITGLGESMESQCGTGVSIHTALK DMLSEITKPSDNYSLSLASRLYAEQTYAILPEYLQCIKELYKESL ETVSFQTAADQARELINSWIESQTNGVIKNFLQPGSVDSQTEL LVNAIYFKGMWEKAFKDEDTQEVPRITEQESRPVQMMYQAG SFKVATVAAEKIKILELPYASGELSMLVLLPDDISGLEQLETTISF EKLTEWTSSNMEDRNMKVYLPRMKIEEKYNLTSVLIALGMT DLFSPAANLSGISAAESLKMSEAIHAAYVEIYEADSEIVSSAGVQ VEVTSDSEEFVRVDHPFLFLIKHNPTNSVLFFGRCISP
Ostrich construct (secretion signal + mature protein)	300	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSdle GDFDVAVLFPSTNNGLLFINTTIIASIAAKEEGVSLEKREAEAG SIGTASAEFCFDVFKELKVHHVNENIFYSPLSIISALSMVYLGAR ENTKTQMEKVIHFDKITGLGESMESQCGTGVSIHTALKDMLSEI TKPSDNYSLSLASRLYAEQTYAILPEYLQCIKELYKESLETVSFQ TAADQARELINSWIESQTNGVIKNFLQPGSVDSQTELVLVNAIY FKGMWEKAFKDEDTQEVPRITEQESRPVQMMYQAGSFKVAT VAAEKIKILELPYASGELSMLVLLPDDISGLEQLETTISFEKLTE WTSSNMEDRNMKVYLPRMKIEEKYNLTSVLIALGMTDLFSP AANLSGISAAESLKMSEAIHAAYVEIYEADSEIVSSAGVQVEVT SDSEEFVRVDHPFLFLIKHNPTNSVLFFGRCISP
Duck OVA sequence as secreted from pichia	301	EAEAGSIGAASTEFCFDVREL RVQHVNENIFYSPFSIISALAMV YLGARDNTRTQIDKVVHFDKLPFGGESMEAQCSTSVSVHSSLR DILTQITKPSDNFSLSFASRLYAEETYAILPEYLQCVKELYKGG LESISFQTAADQARELINSWVESQTNGIKNILQPSSVDSQTTMVL VNAIYFKGMWEKAFKDEDTQAMPFRMTEQESKPQVQMMYQVG SFKVAMVTSEKMKILELPFASGMMSMFVLLPDEVSGLEQLESTI SFEKLTEWTSSSTMMEERRMKVYLPRMKMEEKYNLTSVFMALG MTDLFSSSANMSGISSTVSLKMSEAVHAACVEIFEAGR DVVGS AEAGMDVTSVSEEFRADHPFLFFIKHNPTNSILFFGRWMSP
Duck construct (secretion signal + mature protein)	302	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSdle GDFDVAVLFPSTNNGLLFINTTIIASIAAKEEGVSLEKREAEAG SIGAASTEFCFDVREL RVQHVNENIFYSPFSIISALAMVYLGAR DNTRTQIDKVVHFDKLPFGGESMEAQCSTSVSVHSSLRDILTQI TKPSDNFSLSFASRLYAEETYAILPEYLQCVKELYKGGLESISFQ TAADQARELINSWVESQTNGIKNILQPSSVDSQTTMVLVNAIYF

		KGMWEKAFKDEDTQAMPFRMTEQESKPVQMMYQVGSFKVA MVTSEKMKILELPFASGMMSMFVLLPDEVSGLEQLESTISFEKL TEWTSSTMMEERRMKVYLPRMKMEEKYNLTSVFMALGMTDL FSSSANMSGISSTVSLKMSEAVHAACVEIFEAGR DVVGS AEAG MDVTSVSEEFRA DHPFLFFIKHNPTNSILFFGRWMSP
Ovoglobulin G2	303	TRAPDCGGILTPLGLSYLAEVSKPHA EVVLRQDLMAQRASDLF LGSMEPSRNRITSVKVADLWLSVIPEAGLRLGIEVELRIAPLHAV PMPVRISIRADLHVDMGPDGNLQLLTSACRPTVQAQSTREAES KSSRSILDKVVDVDKLC LDVSKLLLFPNEQLMSLTALFPVTPNC QLQYLPLAAPVFSKQGIALSLQTTTFQVAGAVVPVPVSPVPFSMP ELASTSTSHLILALSEHFYTSLYFTLERAGAFNMTIPSM LTTATL AQKITQVGSLYHEDLPITLSAALRSSPRVLEEGR AALKLFLTV HIGAGSPDFQSFLSVSADV TAGLQLSVSDTRMMISTAVIEDAEL SLAASNVLVRAALLEELFLAPVCQQVPAWMDDVLREGVHLP HLSHFTYTDVNVVVHKDYVLVPCKLKL RSTMA*
Ovoglobulin G3	304	MDSISVTNAKFCFDVF NEMKVHHVNENILYCPLSILTALAMVY LGARGNTESQMKKVLHFDSITGAGSTTDSQCGSSEYVHNLFKE LLSEITRP NATYSLEIADKLYVDKTF SVLPEYLSCARKFYTG GVE EVNFKTAAEEARQLINSWVEKETNGQIKDLLVSSSIDFGTTMVF INTIYFKGIWKIAFNTEDTREMPFSMTKEESKPVQMMCMNNSF NVATLPAEKMKILELPYASGDL SMLVLLPDEVSGLERIEKTINF DKLREWTS TNAMAKKSMKVYLPRMKIEEKYNLT SILMALGMT DLFSRSANLTGISSVDNLMISDAVHG VFMEVNEEGTEATGSTGA IGNIKHSLELEEFRA DHPFLFFIRYNPTNAILFFGRYWSP*
β -ovomucin	305	CSTWGGGHFSTFDKYQYDFTGTCNYIFATV CDESSPDFNIQFRR GLDKKIARI IIELGPSV IIVEKDSISVRSVGVIKLPYASNGIQIAPY GRSVRLVAKLMEMELVVMWNNEDYLMVLTEKKYMGKTCGM CGNYDGYELNDFVSEGKLLD TYKFAALQKMDDPSEICLSEEISI PAIPHKKYAVICSQLLNLVSPTCSVPKDG FVTRCQLDMQDCSEP GQKNCTCSTLSEYSRQCAMSHQVVFNWRTENFCSVGKCSANQI YEECGSPCIKTCSNPEYSCSSHCTYGCFCPEGTVLDDISKNR TCV HLEQCPCTLNGETYAPGDTMKAACRTCKCTMGQWNCKELPCP GRCSLEGGSFVTTFD SRSYRFHGVCTYILMKSSSLPHNGTLMAI YEKSGYSHSETSLSAIIYLSTKDKIVISQNE LLTDDDELKRLPYKS GDITIFKQSSMFIQMHT EFGLELVVQTSPVFQAYVKVSAQFQGR TLGLCGNYNGD TTDDFMTSMDITEGTASLFVDSWRAGNCLPA MERETDPCALSQLNKISAETHCSILTKKGT VFETCHAVVNPTPF YKRCVYQACNYEETFPYICSALGSYARTCSSMGLILENWRNSM DNCTITCTGNQTF SYNTQACERTCLSLSNPTLECHPTDIPIEGCN CPKGMYL NHKNECVRKSHCPCYLED RKYILPDQSTMTGGITCY CVNGRLSCTGKLQNPAESCKAPKKYIS CSDSLENKYGATCAPT CQMLATGIECIPTKCESGCVCADGLYENLDGRCPVPEEC PCEYG GLSYGKGGEQIQTECEICTCRKGKWKCVQKSRCSSSTCNLYGEGHI TTFDGQRFVFDGNCEYILAMDGCNVNRPLSSFKIVTENVICGKS GVTCRSRSISIYLG NLTILRDETY SISGKNLQVKYNVKKNALHLM FDIIIPGKYNMTLIWNKHMNFFIKISRETQETICGLCGNYNGNMK DDFETR SKYVASNELEFVNSWKENPLCGDVYFVVDPCSKNPYR KAWAEKTC SIINSQVFSACHNKVN RMPYYEACVRDSCGCDIGG DCECMCDAIAVYAMACLDKGICIDWRTPEFCPVYCEYYNSHRK TGSGGAYSYGSSVNCTWHYRPCNCPNQYYKYVNI EGCYNCSH

		DEYFDYEKEKCMPCAMQPTSVTLPTATQPTSPSTSSASTVLTET TNPPV*
Lysozyme	306	KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQ ATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLS SDITASVNC AKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRG CRL*
Lysozyme	307	KVFGRCELAAAMKRHGLDNYRGYSLGNWVCVAKFESNFNTQ ATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLS SDITASVNC AKKIVSDGNGMSAWVAWRNRCKGTDVQAWIRG CRL*
Lysozyme C (Human)	308	KVFERCELARTLKLRLGMDGYRGISLANWMCLAKWESGYNTRA TNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSALL QDNIADAVACAKRVVRDPQGIRAWVAWRNRQCNRDVRQYVQ GCGV*
Lysozyme C (Bos taurus)	309	KVFERCELARTLKKLGLDGYKGVSLANWLCLTKWESSYNTKA TNYNPSSESTDYGIFQINSKWWCNDGKTPNAVDGCHVSCRELM ENDIAKAVACAKHIVSEQGITAWVAWKSHCRDHDVSSYVEGC TL*
Ovoinhibitor	310	IEVNCSLYASGIGKDGTSWVACPRNLKPVC GTDGSTYSNECGIC LYNREHGANVEKEYDGECPKHVMIDCSPYLQVVRDGNMVA CPRILKPVCGSDSFTYDNECGICAYNAEHHTNISKLHDGECKLEI GSVDCSKYPSTVSKDGR TLVACPRILSPVCGTDGFTYDNECGIC AHNAEQRTHVSKKHDGKCRQEIPEIDCDQYPTRKTTGGKLLVR CPRILLPVCGTDGFTYDNECGICAHNAQHGTVEKKSHDGRCKE RSTPLDCTQYLSNTQNGEAITACPFILQEVCGTDGVTYSNDCSL CAHNIELGTSVAKKHDGRCREEVPELDCSKYKTSTLKDGRQVV ACTMIYDPVCATNGVTYASECTLCAHNLEQRTNLGKRKNGRC EEDITKEHCREFOKVSPICMEYVPHCGSDGVTYSNRCFFCNAY VQSNRTLNLVSMAAC*
Cystatin	311	MAGARGCVLLAAALMLVGAVLGSEDRSRLLGAPVPVDENDE GLQRALQFAMAEYNRASNDKYSSRVVRVISAKRQLVSGIKYIL QVEIGRTTCPKSSGDLQSCEFHDPEMAKYTTCTFVVYSIPWLN QIKLLESKCQ*
Porcine Lipase	312	SEVCFPRLGCFSDDAPWAGIVQRPLKILPWSPKDVDTRFLLYTN QNQNNYQELVADPSTITNSNFRMDRKTFRFHHGFIDKGEEDWLS NICKNLFKVESVNCICVDWKGGSRGTGTQASQNIRIVGAEVAY FVEVLKSSLGYSPSNVHVIGHSLGSHAAGEAGRRTNGTIERITGL DPAEPCFQGTPELVRLDPSDAKFVDVIHTDAAPIPNLGF GMSQT VGHLDFFPNGGKQMPGCQKNILSQIVDIDGIWEGTRDFVACNH LRSYKYYADSILNPDGFAGFPCDSYNVFTANKCFPCPSEGCPQM GHYADRFPGKTNGVSQVFYLTG DASNFARWRYKVSVTLSGK KVTGHILVSLFGNEGNSRQYEIYKGTLPDNTHSDEFSDSDEVG DLQKV KFIWYNNNVINPTLPRVGASKITVERNDGKVYDFCSQE TVREEVLLTLNPC*
Kid Lipase	313	GLVAADRITGGKDFRDIESKFALRTPEDTAEDTCHLIPGVTESV ANCHFNHSSKTFVVIHGWTVTGMYESWVPKLVAALYKREPDS NVIVVDWLSRAQQHYVPVSAGYTKLVGQDVAKF MNWMADEFN YPLGNVHLLGYSLGAHAAGIAGSLTSKKVN RITGLDPAGPNFE YAEAPSRLSPDDADFVDVLHTFTRGSPGRSIGIQKPVGHVDIYP NGGTFQPGCNIGEALRVIAERGLGDVDQLVKCSHERSVHLFIDS LLNEENPSKAYRCNSKEAFEKGLCLSCRKNRCNNM GYEINKVR

		AKRSSKMYLKTRSQMPYKVFHYQVKIHFSGTESNTYTNQAFEI SLYGTVAESENIPFTLPEVSTNKTYSFLLYTEVDIGELLMLKLK WISDSYFSWSNWWSSPGFDIGKIRVKAGETQKKVIFCSREKMS YLQKGKSPVIFVKCHDKSLNRKSG*
Porcine Lactoferrin	314	APKKGVRWCVISTA EYSKCRQWQSKIRRTNPMFCIRRASPTDCI RAIAAKRADAVTLDGGLVFEADQYKLRPVAAEIYGTEENPQTY YYAVAVVKKGFNFQLNQLQGRKSCHTGLGRSAGWNIPIGLLR FLDWAGPPEPLQKAVAKFFSQSCVPCADGNAYPNLCQLCIGKG KDKCACSSQEPYFGYSGAFNCLHKGIGDVAFVKESTVFENLPQ KADRDKYELLCPDNTRKPVEAFRECHLARVPSHAVVARSVNG KENSIWELLYQSQKKFGKSNPQEFQLFGSPGQQKDLLFRDATIG FLKIPSKIDSKLYLGLPYLTAIQGLRETA AEVEARQAKVWCAV GPEELRKCRQWSSQSSQNLNCSLASTTEDCIVQVLKGEADAMS LDGGFIYTAGKCGLVPVLAENQKSRQSSSSDCVHRPTQGYFAV AVVRKANGGITWNSVRGTSCHTAVDRTAGWNIPMGLLVNQT GSCKFDEFFSQSCAPGSQPGSNLCALCVGNDQGVDKCVPNSNE RYYGYTGAFRCLAENAGDVA FVKDVTVLDNTNGQNTTEWAR ELRSDDFELLCLDGTRKPVTEAQNCHLAVAPSHAVVSRKEKAA QVEQVLLTEQAQFGRYGKDCPDKFCLFRSETKNLLFNDNTEVL AQLQGKTTYEKEYLGSEYVTAIANLKQCSVSPLLEACAFMMR*
Bovine Lactoferrin	315	APRKNVRWCTISQPEWFKCRRWQWRMKKLGAPSITCVRRAFA LECIRAI AEKKADAVTLDGGMVFEAGRDPYKLRPVAAEIYGTK ESPQTHYYAVAVVKKGSNFQLDQLQGRKSCHTGLGRSAGWIIP MGILRPYLSWTESLEPLQGAVAKFFSASCVCIDRQAYPNLCQL CKGEGENQCACSSREPYFGYSGAFKCLQDGAGDVA FVKETTTF ENLPEKADRDQYELLCLNNSRAPVDAFKECHLAQVP SHAVVA RSVDGKEDLIWKLLSKAQEKFGKNKSRSFQLFGSPPGQRDLLF KDSALGFLRIPSKVDSALYLGSRYLTTLKNLRETAEEVKARYTR VWCAVGPEEQKKCQQWSQQSGQNVTCATASTDDCIVLVLK GEADALNLDGGYIYTAGKCGLVPVLAENRKSSKHSSSLDCVLRP TEGYLAVAVVKKANEGLTWNSLKD KKSCHTAVDRTAGWNIP MGLIVNQTGSCAFDEFFSQSCAPGADPKSRLCALCAGDDQGLD KCVPSKEKYYGYTGAFRCLAEDVG DVA FVKNDTVWENTNG ESTADWAKNLNREDFRLLCLDGTRKPVTEAQSCHLAVAPNHA VVSRS DRAAHVKQVLLHQALFGKNGKNCPDKFCLFKSETKN LLFNDNTECLA KLGGRPITYEYLGTEYVTAIANLKKCSTSPLLE ACAFLTR*

Table 6: Exemplary Linkers

Sequence Info	SEQ ID NO	Amino Acid sequence
GGGS linker	SEQ ID NO : 316	GGGGS

GSS linker	SEQ ID NO : 317	GSS
A rigid linker that forms 4 turns of an alpha helix	SEQ ID NO : 318	EAAAREAAAREAAAREAAAR
Full linker	SEQ ID NO : 319	GSSGSSGSSGSSGSSGSSGSSGSSEAAAREAAAREAAAREAAARGGG GSGGGGSGGGGS
A flexible GS linker with higher S content	SEQ ID NO : 320	GSSGSSGSSGSSGSSGSSGSSGSS
A flexible GS linker with much higher G content	SEQ ID NO : 321	GGGGSGGGGSGGGGS

Table 7: ALG/OST Pathway knockouts

Sequence Info	SEQ ID NO:	Amino Acid sequence
ALG6 (GS115 - GQ68_00 786T0/XP_0024914 63.1)	SEQ ID NO: 322	MPHKRTPSSSLLYARIPGISFENSPVDFLSPFGPAPNQWVARYIII FAILIRLAVGLGSYSGFNTPPMYGDFAQRHWMEITQHLSIEKWY FYDLQYWGLDYPPLTAFHSYFFGKLGSFINPAWFALDVSRGFESV DLKSYMRTAILSELLCFIPAVIWYCRWMGLNYFNQNAIEQTIIAS AILFNPSLIIDHGHFQYNSVMLGFALLSILNLLYDNFALAAIFFVLS ISFKQMALYYSPIMFFYMLSVSCWPLKNFNLLRLATISIAVLLTFA TLLLFPVLVDGMSQIGQILFRVFPFSRGLFEDKVANFWCTTNILVK YKQLFTDKTLTRISLVATLIAISPSCFIIFTHPKKVLLPWAFACSW AFYLFQVHEKSVLVPLMPTLLLVKDLDIISMVCWISNIAFFS MWPLLKRDGLALEYFVLGILSNWLIGNLNWISKWLVPSFLIPGPT

		LSKKVPRDRTKTVVHTHWFWGSVTFVSYLGATVIQFVDWLYLPP AKYPDLWVILNTTSLFACFGLFWLWINYNLYILRDFKLKDA*
STT3 (GS115 - Q68_0166 9T0/XP_0 02490630. 1)	SEQ ID NO: 323	MVTINDQGYITVNDRLKLIKSLILVIFISITIAAVSSRLFSVIRFESI IHEFDPWFNFRATKYL VHNGFYKFLNWFDDBKWTYPLGRVTGGTL YPGLMVTSAVIHNLAKIGLPIDIRNICVMLAPAFSSSLTAIAMYFLT LELTNDSESIANGTAKATAALFSAIFMGITPGYISRSVAGSYDNEAI AITLLMVTIFYFWIKAVKLGSIIFYSSVTALFYFYMVSAWGGYVFIT NLIPLHVFVLLMGRFTHKIYVSYTTWYVLGTLMSMQIPFVGFLPI RSNDHMAPLGVFGLIQLVLIGDFFKSQLSRKVFIKLAIASGVVIGIL GVVGLVLATKIGLIAPWTGRFYSLWDTNYAKIHIPIIASVSEHQPTP WASFFFDLNFILWLPVGVWFCFQELTDGAVFVIISVLASYFAG VMVRLILTAPIVCVCGAIAITKLFEVYSDFTDVVKGKSGNFFTLF SKLAVLGSFGFYLFYVVKHCTWVTENAYSSPSVVLASHAADGSQI LIDDYREAYWLRMNTPEDAKVMAWWDYGYQIGGMADRTTFV DNNTWNNTHIATVGKAMAVSEEKSEVIMRQLGVVDYILVIFGGVL GYSGDDINKFLWMVRISSEGIWPEEVSERGYFTPRGEYKIDDNAAQ AMKDSMLYKMSFYRFGELFPSGDAIDRVRGQRLRSYAESIDLNI VEEVFTSENWLVRLYKLKEPDNLGRSLTLKDNEKKLATKKGRR LRVNKKPSLDLRV*

EXAMPLES

[0222] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1: Expression Constructs, transformation, protein purification and processing

[0223] Constructs may be designed to disrupt beta-mannosyl transferases BMT1 and BMT2 genes (XP_002493882.1 and XP_002493883.1 respectively). Additionally, expression constructs may be designed to express one or more proteins of interest, such as nutritional proteins. The constructs may be transformed into a host cell such as *Pichia pastoris*.

[0224] In one example, another expression construct expressing a mannosidase may be designed and transformed into the host cell. In this example, the disruption of BMT1 and BMT2 would lead to the production of a smaller exopolysaccharide. Additionally, the mannosidase production would be expected to further hydrolyze the exopolysaccharide to mannose which can be used by the host cell as a carbon source. It would be expected that the host cell produces a reduced level of exopolysaccharides thereby reducing the impurities to be separated from the recombinantly produced nutritional protein.

[0225] The nutritional protein may be secreted from the host cell and purified using conventional methods of purification.

Example 2: Expression Constructs, transformation, protein purification and processing

[0226] Constructs were designed to disrupt beta-mannosyl transferases BMT1 and BMT2 genes (XP_002493882.1 and XP_002493883.1 respectively) in a *Pichia pastoris* strain. Knockouts were performed via standard Homologous Recombination (HR) methods in yeast. In summary, genes of interest (GOIs) were deleted by using linearized plasmids that had homology to genomic regions that surround the GOIs, which were transformed into yeast via standard electroporation techniques. The native HR machinery replaces the GOI with the linearized plasmid. The plasmid with antibiotic resistance can eventually be removed using the Cre/lox recombinase system leaving only a small insertion scar where the GOI initially was found.

[0227] In this example, the disruption of BMT1 and BMT2 lead to the production of a smaller exopolysaccharide. Using gel electrophoresis and the cationic dye Alcian blue (which binds to the phospho-mannan moiety via the phosphodiester bond) it is shown in FIG. 1 that disrupting the *BMT1* and *BMT2* genes (AT250_GQ6804781 and AT250_GQ6804782) produces a noticeable shift in the size of EPS, which strongly suggests that the EPS byproduct is a form of mannan polysaccharide.

[0228] It is also shown in FIG. 2 that *Pichia* species can grow with mannose as a sole carbon source, illustrating that production strains will be able to recover carbon from the EPS/mannan that is broken down.

Example 3: Expression Constructs, transformation, protein purification and processing

[0229] Several *Pichia pastoris* strains which were previously transformed to express a glycoprotein (ovomucoid) and a transcription factor (HAC1) were cultured. The supernatant from that culture contained exopolysaccharides (EPS). The EPS was filter-purified and analyzed. Additionally, Strain 1 and Strain 2 were transformed with a mannosidase expressing constructs (pPMP20 SDBT2623-2631 vs pTKL3 SDBT2623). The EPS produced by these strains were analyzed and as is shown in FIG. 3, the size of the EPS byproduct is unchanged when strains are incubated with purified EPS. The Sed1 display construct found in the strain uses the PMP20 promoter from *Pichia pastoris* and TDH3 terminator.

[0230] The cells were also incubated with their own culture supernatant to see if increasing the time spent with substrate would allow for hydrolysis of the polysaccharide byproduct. FIG. 4 shows that regardless of the expressed mannosidase (pPMP20 SDBT2623-2631 vs pTKL3 SDBT2623), there is no activity for the enzymes against the wild-type mannan, which is highly branched and ends in terminal beta anomers of mannose.

[0231] While the mannosidases were not able to act on the “wild-type” EPS produced in Strain 1 cells or the purified product, FIG. 5 shows that when the enzymes are coupled with mannosyltransferase deletions, they do indeed use EPS as a substrate. Strain 2 has had the genes responsible for producing terminal beta mannose anomers (BMT1 and BMT2, GQ6804782 and GQ6804781, respectively), and an alpha-1,2 branching enzyme (MNN2 family protein, GQ6802166), which already produces a right shift in the elution profile of the EPS it produces. When this deletion mutant is coupled with the expression of different mannosidase constructs, it produces a right shift in the elution time of the EPS byproduct, suggesting that the enzymes display activity against the simplified structure of mannan following the deletion of native mannan mannosyltransferases.

Example 4: Surface display of mannosidases

[0232] Mannan has been identified using gel electrophoresis and mass spectrometry as the polysaccharide impurity (known as EPS – extracellular polysaccharide) found in supernatants from *P. pastoris* strains that secrete Proteins of Interest (POIs). Mannan is produced by the sequential action of many mannosyltransferases in the Golgi apparatus. Following the attachment of the core glycan moiety to an asparagine residue, mannan polymerase I (M-pol I) extend the core structure with ~10 alpha-1,6 mannose units using the Mnn9 catalytic subunit. Next the M-pol II complex (catalytic subunits Mnn10 and Mnn11) extends by another ~50-100 alpha-1,6 mannose units, which creates a long, linear mannan backbone composed of alpha-1,6-linked sugars. The linear mannan backbone is the extensively decorated with alpha-1,2- and phospho-mannose branch points. These decorations are carried out by members of the MNN and KTR families of proteins – of which there are a total of 10 known in *P. pastoris*. Finally, some species of yeast (including *C. albicans* and *P. pastoris*) produce terminal beta-1,2-linked mannose units to “cap” the mannan molecule (opposed to the terminal alpha-1,3-mannose units found in *S. cerevisiae* mannan), and these reactions are carried out by the BMT family of mannosyltransferases (four of these family members are found in *P. pastoris*, two of which have been determined to be catalytically active – BMT1/2). Following the identification of the mannosyltransferases discussed in Example 2, they were deleted to reduce the size and

complexity of the mannan/EPS molecule. As is shown in the chromatogram in FIG. 6, the deletion of multiple native mannosyltransferases indeed increased the retention time of eluted EPS using size exclusion chromatography (SEC) (indicative of a decrease in the size of the molecule). Strain 3 was built from Strain 1 by the sequential deletion of five native mannosyltransferases (BMT1 (SEQ ID NO: 12), BMT2 (SEQ ID NO: 13), MNN2 (SEQ ID NO: 1), MNNF1 (SEQ ID NO: 2), MNNF2 (SEQ ID NO: 3)), causing the noticeable right-shift in the EPS peak between 8 and 9 minutes.

[0233] The strain was also modified to express mannan hydrolytic enzymes (mannanases/mannosidases) which are normally expressed by the common human gut microbe *Bacteroides thetaiotaomicron*. Most yeasts are not known to produce enzymes that breakdown their own cell wall material, however *B. theta* has been shown to scavenge carbon in the form of mannose from yeast cell wall material in the human gut. Using a surface-display approach (FIG. 7) this example demonstrates that these enzymes can be used to breakdown the EPS molecule produced by *P. Pastoris* (following the deletion of select native mannosyltransferases), once again evidenced by shifts in the elution profile of EPS following SEC analysis (FIG. 8).

[0234] Some mannosyltransferase deletions are required for *B. theta* mannosidases to recognize EPS as a substrate for cleavage. In FIG. 9, it is shown that when Strain 1 and Strain 2 (Strain 1 + 3 deleted mannosyltransferases) express the exact same mannosidase construct, only the Strain 2+mannosidase build produces EPS which the surface-displayed enzyme can use as a substrate. The disruption of native mannosyltransferases is important for *B. theta* enzymes to recognize mannan as a substrate for cleavage. Only the strain with deletions and mannosidase elicits the right-shift in the EPS elution profile.

CLAIMS

WHAT IS CLAIMED IS:

1. A recombinant host cell for manufacturing a heterologous protein of interest, wherein the host cell is a yeast and is engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof wherein the underexpression is compared to the host cell prior to genetic manipulation to achieve underexpression, wherein the host cell is engineered to express a heterologous protein of interest and a heterologous mannosidase.
2. The recombinant host cell of claim 1, wherein underexpression is achieved by independently for each mannosyl transferase protein knocking-out the polynucleotide encoding the mannosyl transferase protein or a homologue thereof from the genome of said host cell, disrupting the polynucleotide encoding the mannosyl transferase protein or a homologue thereof in the host cell, disrupting a promoter which is operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof, replacing the promoter which is operably linked with said polynucleotide encoding the mannosyl transferase protein or a homologue thereof with another promoter which has lower promoter activity, or disrupting expression control sequences of the mannosyl transferase protein or a homologue thereof, wherein the functional homologue has at least 70% sequence identity to an amino acid sequence of a mannosyl transferase.
3. The recombinant host cell of claim 1 or claim 2, wherein the host cell is *Pichia pastoris*.
4. The recombinant host cell of any one of the preceding claims, wherein the BMT1 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 12.
5. The recombinant host cell of any one of the preceding claims, wherein the BMT2 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 13.
6. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to express at least 10% less BMT1 relative to a host cell which has not been engineered to underexpress BMT1.
7. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%,

90%, or 95% less BMT1 relative to a host cell which has not been engineered to underexpress BMT1.

8. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to knock out BMT1, wherein the knockout leads to no activity of BMT1 in the recombinant host cell.

9. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to express at least 10% less BMT2 relative to a host cell which has not been engineered to underexpress BMT2.

10. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less BMT2 relative to a host cell which has not been engineered to underexpress BMT2.

11. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is engineered to knock out BMT2, wherein the knockout leads to no activity of BMT2 in the recombinant host cell.

12. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell produces a reduced size of exopolysaccharides relative to a host cell not engineered to underexpress BMT1 and BMT2.

13. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is further engineered to underexpress alpha-1,2-mannosyltransferase MNN2.

14. The recombinant host cell of claim 13, wherein the MNN2 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 1.

15. The recombinant host cell of claim 13, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNN2 relative to a host cell which has not been engineered to underexpress MNN2.

16. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is further engineered to underexpress MNMF1.

17. The recombinant host cell of claim 16, wherein the MNMF1 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 2.

18. The recombinant host cell of claim 16, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNMF1 relative to a host cell which has not been engineered to underexpress MNMF1.

19. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is further engineered to underexpress MNNF2.
20. The recombinant host cell of claim 19, wherein the MNNF2 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 3.
21. The recombinant host cell of claim 19, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less MNNF2 relative to a host cell which has not been engineered to underexpress MNNF2.
22. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell is further engineered to underexpress one or more enzymes in addition to BMT1 and BMT2.
23. The recombinant host cell of claim 22, wherein the one or more enzyme comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 4-11, 14-15, and 72-85.
24. The recombinant host cell of claim 22, wherein the recombinant host cell is engineered to express at least 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% less one or more enzymes relative to a host cell which has not been engineered to underexpress said one or more enzymes.
25. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell recombinantly expresses a mannosidase from a species different from the recombinant host cell.
26. The recombinant host cell of claim 25, wherein the mannosidase is from a genus different from the recombinant host cell.
27. The recombinant host cell of claim 25 or claim 26, wherein the mannosidase comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 41-56.
28. The recombinant host cell of any one of claims 25-27, wherein the mannosidase is expressed on the surface of the recombinant host cell.
29. The recombinant host cell of any one of claims 25-28, wherein the recombinant host cell expresses a surface-displayed fusion protein comprising a catalytic domain of a mannosidase and an anchoring domain of a glycosylphosphatidylinositol (GPI)-anchored protein, wherein the

anchoring domain comprises at least about 200 amino acids and/or at least about 30% of the residues in the anchoring domain are serines or threonines.

30. The recombinant host cell claim 29, wherein the anchoring domain comprises at least about 225 amino acids, at least about 250 amino acids, at least about 275 amino acids, at least about 300 amino acids, at least about 325 amino acids, at least about 350 amino acids, at least about 375 amino acids, or at least about 400 amino acids.

31. The recombinant host cell of claim 29 or claim 30, wherein at least about 35% of the residues in the anchoring domain are serines or threonines, at least about 40% of the residues in the anchoring domain are serines or threonines, at least about 45% of the residues in the anchoring domain are serines or threonines, or at least about 50% of the residues in the anchoring domain are serines or threonines.

32. The recombinant host cell of any one of claims 29-31, wherein the serines or threonines in the anchoring domain are capable of being O-mannosylated.

33. The recombinant host cell of any one of claims 29-32, wherein a fusion protein having an anchoring domain comprising at least about 325 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain comprising less than about 300 amino acids.

34. The recombinant host cell of any one of claims 29-33, wherein a fusion protein having an anchoring domain comprising at least about 300 amino acids provides greater enzymatic activity relative to a fusion protein having an anchoring domain comprising less than about 250 amino acids.

35. The recombinant host cell of any one of claims 29-34, wherein the fusion protein comprises the anchoring domain of the GPI anchored protein.

36. The recombinant host cell of any one of claims 29-35, wherein the fusion protein comprises the GPI anchored protein without its native signal peptide.

37. The recombinant host cell of any one of claims 29-36, wherein the GPI anchored protein is not native to the recombinant host cell.

38. The recombinant host cell of any one of claims 29-37, wherein the GPI anchored protein is naturally expressed by a *S. cerevisiae* cell and the recombinant host cell is not a *S. cerevisiae* cell.

39. The recombinant host cell of any one of claims 29-38, wherein the GPI anchored protein is selected from Tir4, Dan1, Dan4, Sag1, Fig2, and Sed1.

40. The recombinant host cell of any one of claims 29-39, wherein the anchoring domain of the GPI anchored protein comprises an amino acid sequence that is at least 70% identical, at least

75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NO: 57 to SEQ ID NO: 71.

41. The recombinant host cell of any one of claims 29-40, wherein the anchoring domain of the GPI anchored protein comprises an amino acid sequence of one of SEQ ID NO: 57 to SEQ ID NO: 71.

42. The recombinant host cell of any one of claims 29-41, wherein the recombinant host cell comprises a genomic modification that expresses the fusion protein and/or comprises an extrachromosomal modification that expresses the fusion protein.

43. The recombinant host cell of any one of claims 29-42, wherein the fusion protein comprises a portion of the mannosidase in addition to its catalytic domain.

44. The recombinant host cell of any one of claims 29-43, wherein the fusion protein comprises substantially the entire amino acid sequence of the mannosidase.

45. The recombinant host cell of any one of claims 29-44, wherein in the fusion protein, the catalytic domain is N-terminal to the anchoring domain.

46. The recombinant host cell of any one of claims 29-45, wherein the fusion protein comprises a linker between the catalytic domain and the anchoring domain.

47. The recombinant host cell of any one of claims 29-46, wherein the fusion protein comprises a linker having an amino acid sequence that is at least 95% identical to any one of SEQ ID NOS: 316-321.

48. The recombinant host cell of any one of claims 29-47, wherein, upon translation, the fusion protein comprises a signal peptide and/or a secretory signal.

49. The recombinant host cell of any one of claims 29-48, wherein the recombinant host cell comprises two or more fusion proteins, three or more fusion proteins, or four fusion proteins.

50. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell comprises a mutation in its AOX1 gene and/or its AOX2 gene.

51. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell comprises a genomic modification that overexpresses a secreted heterologous protein of interest and/or comprises an extrachromosomal modification that overexpresses a secreted protein of interest.

52. The recombinant host cell of any one of the preceding claims, wherein the secreted protein of interest is an animal protein.

53. The recombinant host cell of claim 52, wherein the animal protein is an egg protein.

54. The recombinant host cell of claim 53, wherein the egg protein is selected from the group consisting of ovalbumin, ovomucoid, lysozyme ovoglobulin G2, ovoglobulin G3, α -ovomucin,

β -ovomucin, ovotransferrin, ovoinhibitor, ovoglycoprotein, flavoprotein, ovomacroglobulin, ovostatin, cystatin, avidin, ovalbumin related protein X, and ovalbumin related protein Y.

55. The recombinant host cell of any one of claims 52 to 54, wherein the genomic modification and/or the extrachromosomal modification that overexpresses the secreted recombinant protein comprises an inducible promoter.

56. The recombinant host cell of claim 55, wherein the inducible promoter is an AOX1, DAK2, PEX11, FLD1, FGH1, DAS1, DAS2, CAT1, MDH3, HAC1, BiP, RAD30, RVS161-2, MPP10, THP3, TLR, GBP2, PMP20, SHB17, PEX8, PEX4, or TKL3 promoter.

57. The recombinant host cell of any one of claims 52 to 56, wherein the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein comprises an AOX1, TDH3, MOX, RPS25A, or RPL2A terminator.

58. The recombinant host cell of any one of claims 52 to 57, wherein the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein encodes a signal peptide and/or a secretory signal.

59. The recombinant host cell of any one of claims 52 to 58, wherein the genomic modification and/or the extrachromosomal modification that overexpresses a secreted recombinant protein comprises codons that are optimized for the species of the recombinant host cell.

60. The recombinant host cell of any one of claims 52 to 59, wherein the secreted recombinant protein is designed to be secreted from the cell and/or is capable of being secreted from the cell.

61. The recombinant host cell of any one of claims 56 to 60, wherein the additional genomic modification reduces the number of native cell wall proteins expressed by the recombinant host cell, thereby allowing additional space for localization of the surface-displayed fusion protein.

62. The recombinant host cell of any one of the preceding claims, wherein the recombinant host cell comprises a further genomic modification that overexpresses a protein related to the p24 complex.

63. The recombinant host cell of claim 62, wherein the recombinant host cell comprises a further genomic modification comprising that overexpresses more than one protein related to the p24 complex.

64. The recombinant host cell of claim 62 or claim 63, wherein the protein related to the p24 complex is selected from Erp1, Erp2, Erp3, Erp5, Emp24, and Erv25.

65. The recombinant host cell of any one of claims 62 to 64, wherein the protein related to the p24 complex comprises the amino acid sequence of any one of SEQ ID NO: 86 to SEQ ID NO: 91.

66. A method for expressing a heterologous protein of interest, the method comprising obtaining a recombinant host cell of any one of claims 1 to 65 and culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest.
67. An isolated heterologous protein of interest expressed according to the method of claim 66.
68. Use of the isolated heterologous protein of interest of claim 67 in the manufacture of a nutritional, dietary, digestive, supplements, such as in food products, feed products, or cosmetic products.
69. A method for expressing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising obtaining a recombinant host cell of any one of claims 1 to 65 and culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest.
70. An isolated heterologous protein of interest expressed according to the method of claim 69.
71. Use of the isolated heterologous protein of interest of claim 70 in the manufacture of a nutritional, dietary, digestive, supplements, such as in food products, feed products, or cosmetic products.
72. A method for expressing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising:
obtaining a host cell that is a yeast and is engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof wherein the underexpression is compared to the host cell prior to genetic manipulation, wherein the host cell is engineered to express a heterologous protein of interest and a heterologous mannosidase; and
culturing the recombinant host cell under conditions that allow expression of the heterologous protein of interest
73. The method of claim 72, wherein the BMT1 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 12 and the BMT2 protein comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to SEQ ID NO: 13.

74. The method of claim 72 or claim 73, wherein the recombinant host cell is further engineered to underexpress one or more enzymes comprising an amino acid sequence of one of SEQ ID NOs: 1-11, 14-15, and 72-85.
75. The method of any one of claims 72 to 74, wherein the recombinant host cell recombinantly expresses a mannosidase from a species different than from the recombinant host cell.
76. The method of claim 75, wherein the mannosidase comprises an amino acid sequence that is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to one of SEQ ID NOs: 41-56.
77. The method of claim 75 or claim 76, wherein the mannosidase is expressed on the surface of the recombinant host cell.
78. The method of any one of claims 72 to 77, wherein the recombinant host cell expresses a surface-displayed fusion protein comprising a catalytic domain of a mannosidase and an anchoring domain of a glycosylphosphatidylinositol (GPI)-anchored protein, wherein the anchoring domain comprises at least about 200 amino acids and/or at least about 30% of the residues in the anchoring domain are serines or threonines.
79. The method of any one of claims 72 to 78, wherein the heterologous protein of interest is secreted from the recombinant host cell.
80. The method of claim 79, wherein the secreted heterologous protein of interest is an animal protein.
81. The method of claim 80, wherein the animal protein is an egg protein.
82. The method of claim 81, wherein the egg protein is selected from the group consisting of ovalbumin, ovomucoid, lysozyme ovoglobulin G2, ovoglobulin G3, α -ovomucin, β -ovomucin, ovotransferrin, ovoinhibitor, ovoglycoprotein, flavoprotein, ovomacroglobulin, ovostatin, cystatin, avidin, ovalbumin related protein X, and ovalbumin related protein Y.
83. The method of any one of claim 72 to 82, wherein the recombinant host cell comprises a further genomic modification that overexpresses a protein related to the p24 complex.
84. An isolated heterologous protein of interest expressed according to the method of any one of claims 72 to 83.
85. Use of the isolated heterologous protein of interest of claim 84 in the manufacture of a nutritional, dietary, digestive, supplements, such as in food products, feed products, or cosmetic products.
86. A method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising:

obtaining a yeast cell engineered to express a heterologous protein of interest and/or a heterologous mannosidase; and
modifying the yeast cell to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof. A method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising:

obtaining a yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous mannosidase; and
modifying the yeast cell to express a heterologous protein of interest.

87. A method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising:

obtaining a yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous protein of interest; and
modifying the yeast cell to express a heterologous mannosidase.

88. A method for manufacturing a recombinant host cell for manufacturing a heterologous protein of interest having of a reduced level of exopolysaccharides, the method comprising:

obtaining a yeast cell
modifying the yeast cell engineered to underexpress two mannosyl transferases: beta-mannosyl transferase 1 (BMT1) and beta-mannosyl transferase 2 (BMT2) or functional homologues thereof and engineered to express a heterologous protein of interest;
modifying the yeast cell to express a heterologous protein of interest; and
modifying the yeast cell to express a heterologous mannosidase.

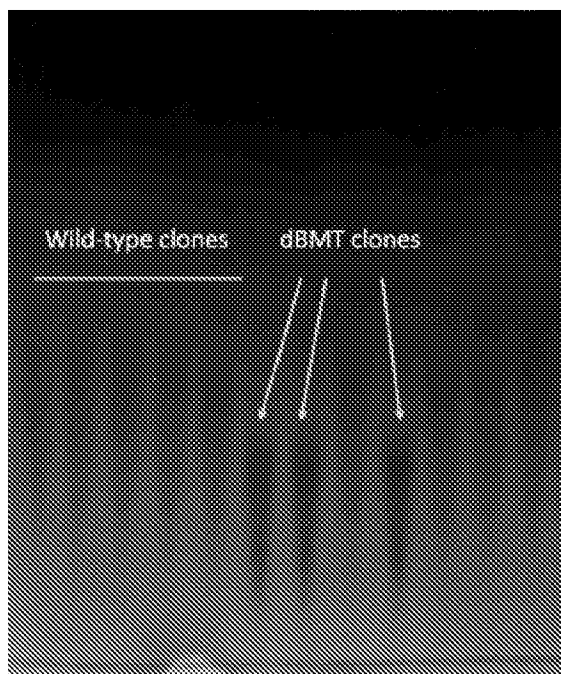


FIG. 1

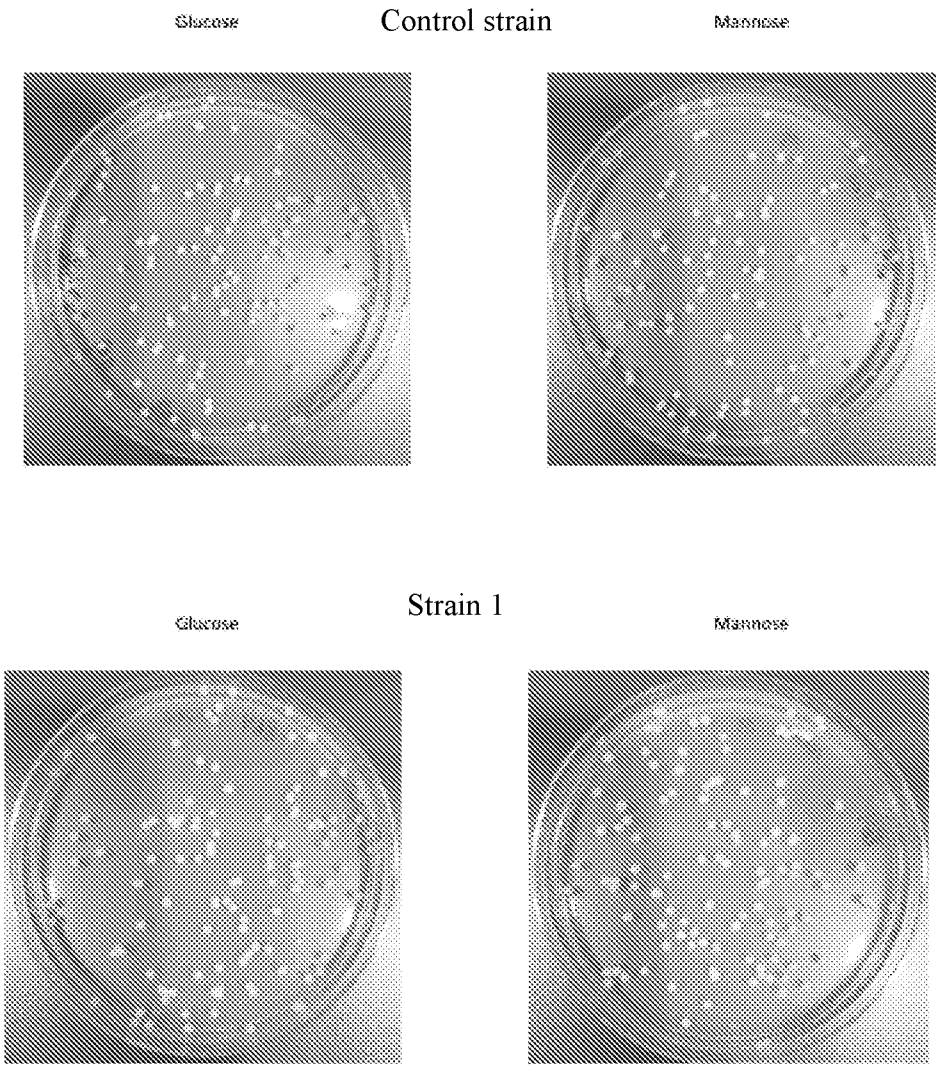


FIG. 2

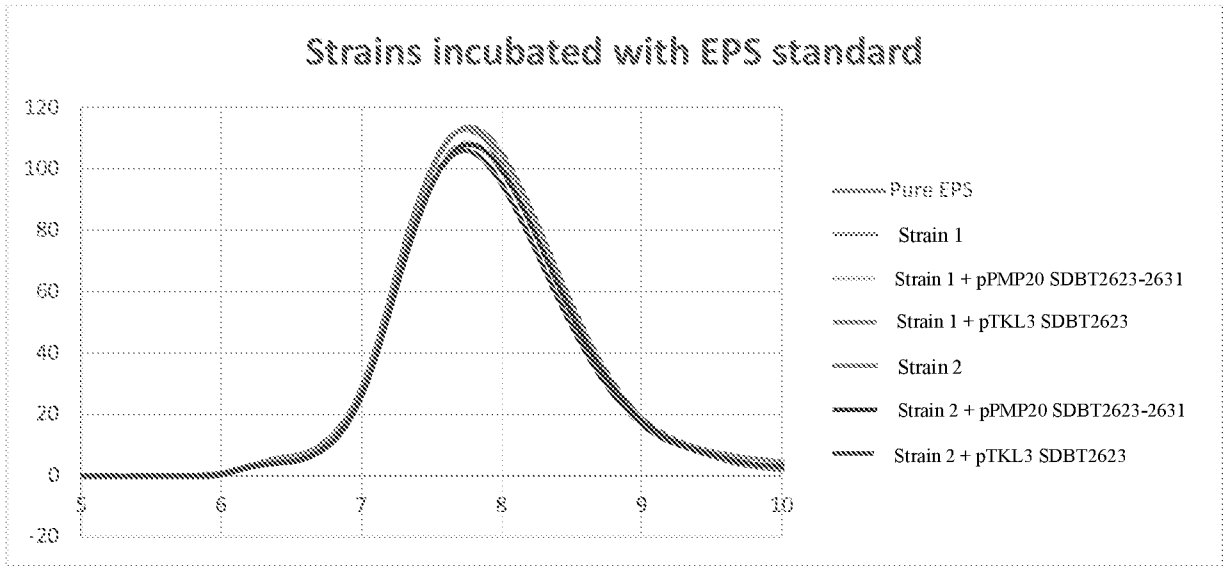


FIG. 3

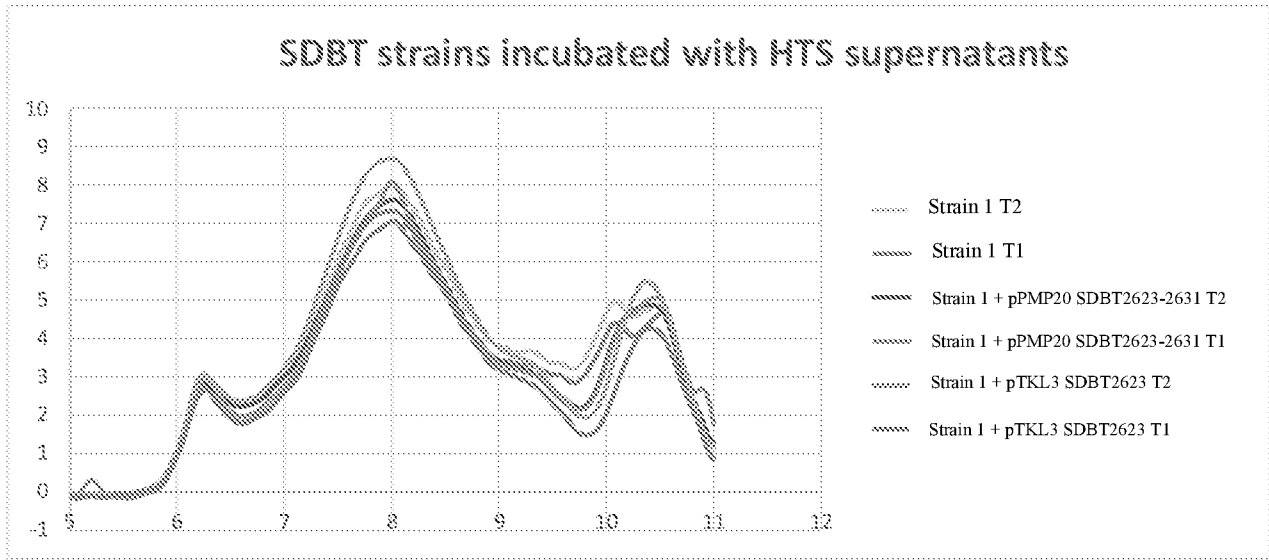


FIG. 4

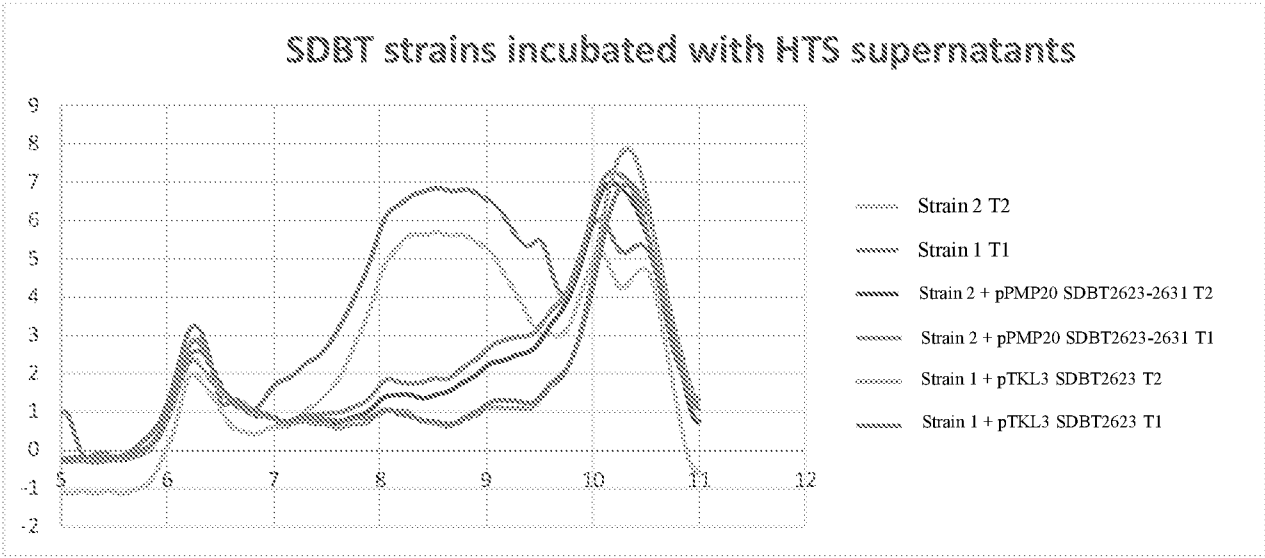


FIG. 5

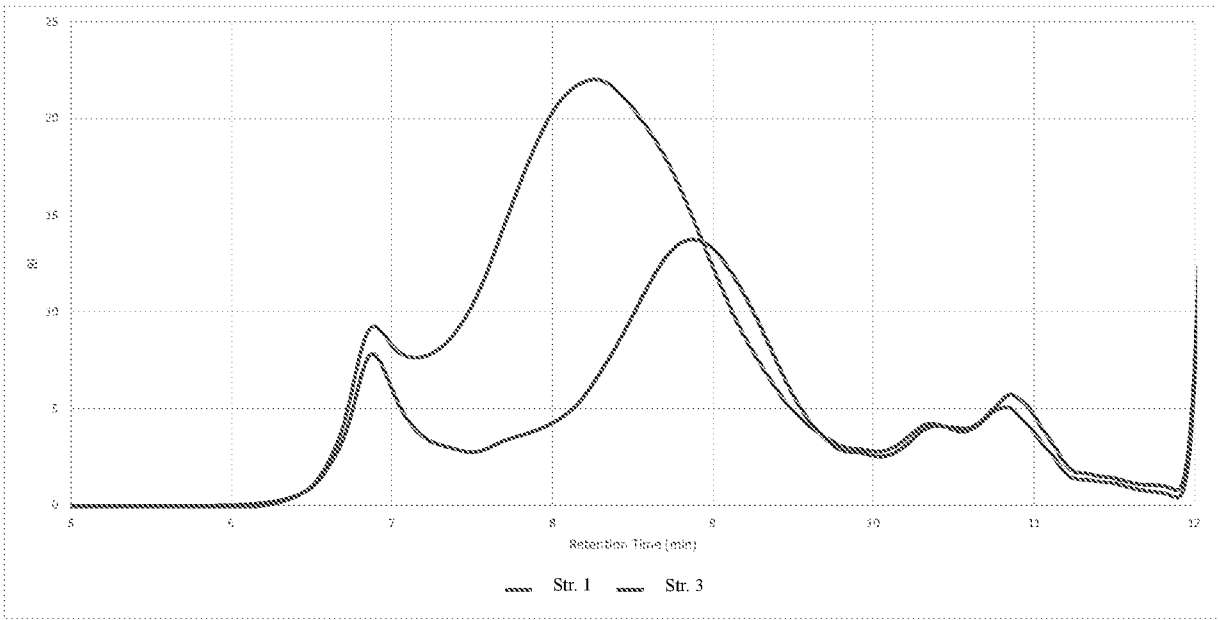


FIG. 6



FIG. 7

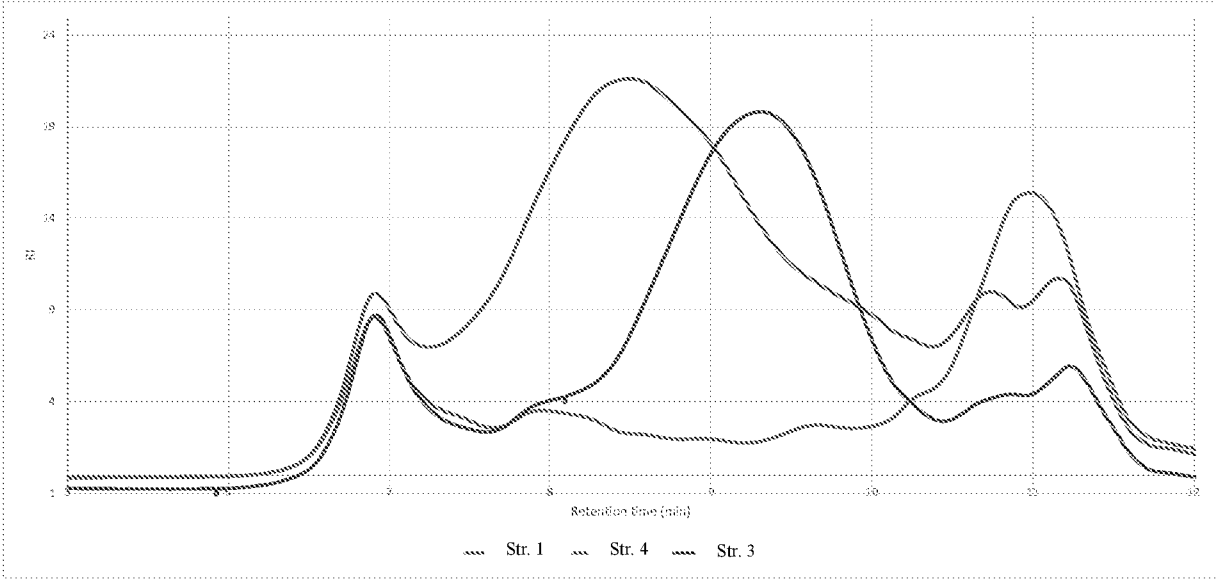


FIG. 8

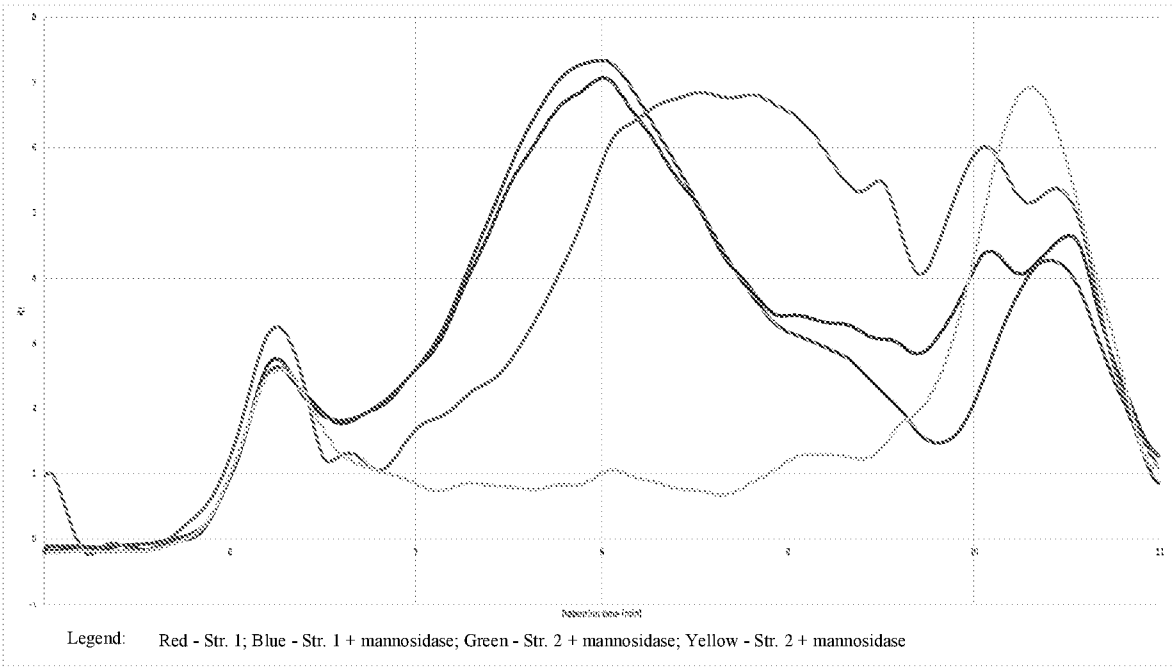


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/038095

A. CLASSIFICATION OF SUBJECT MATTER

C12N 1/16 (2006.01) C12N 15/81 (2006.01) C12R 1/84 (2006.01) C07K 14/395 (2006.01) C07K 14/81 (2006.01)
A23J 1/18 (2006.01) A23L 33/195 (2016.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

REGISTRY, BIOSIS, CAPLUS, EMBASE, MEDLINE(β -1,2-mannosyltransferase, under expression β -1,2-mannosyltransferase, reduced exopolysaccharide, yeast, heterologous mannosidase, under expression MNN2; and other synonyms and like terms); PATENW (C12N/16, C12N15/81, and same keywords, synonyms and like terms as above); GENOMEQUEST (SEQ ID NO:1, 12 and 13); ESPACENET, PUBMED (Applicant and Inventor Search).

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	



Further documents are listed in the continuation of Box C



See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"D" document cited by the applicant in the international application	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 13 October 2022	Date of mailing of the international search report 13 October 2022
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustalia.gov.au	Authorised officer Tycho Neumann AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61 2 6283 3169

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/US2022/038095
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/115903 A2 (MERCK SHARP & DOHME CORP.) 30 August 2012 Abstract; Examples 6 and 7; Figure 1; Claims 1, 5, 17 and 18; par.[0074-0076]; D7(Example 3; pg.13 line 36 to pg.14 line 29)	1-88
X	HOPKINS, D. et al., 'Elimination of β -mannose glycan structures in <i>Pichia pastoris</i> ', 2011. Glycobiology, Vol. 21, No. 12, pages 161-1626 Title; Abstract; Materials and Methods: Generation of BMT gene knock-out strains; Materials and Methods: Reagents and strains; Table 2; Figures 4-6; D8(Whole Document)	1-88
X	WO 2011/106389 A1 (MERCK SHARP & DOHME CORP.) 01 September 2011 Claim 1; pg.11 lines 25-34; pg.12 lines 11-13; pg.12 lines 26-30; pg.43 lines 6-13; pg.47 lines 12-16; Example 5; Figures 1a-1H, 10-15, and 23A-23D.	1-88
X	RABERT, C. et al., 'Recombinant proteins for industrial uses: utilization of <i>Pichia Pastoris</i> expression system', 2013. Brazilian Journal of Microbiology, Vol.44, No.2, pages 351-356. Whole Document	67, 68, 70, 71, 84, 85
X A	JUTURU, V. & WU, J.C., 'Heterologous Protein Expression in <i>Pichia Pastoris</i> : Latest Research Progress and Applications', 2018. ChemBioChem, Vol.19, No.1, pages 7-21. Section 13. Production of Cellulases and Hemicellulases Whole Document	67, 68, 70, 71, 84, 85 1-88
X A	AHMAD, M. et al., 'Protein expression in <i>Pichia pastoris</i> : recent achievements and perspectives for heterologous protein production', 2014. Appl Microbiol Biotechnol, Vol.98, No.12, pages 5301-5317. par. spanning pg. 5301 R col. to pg.5302 L col.; pg.5311 R Col. 2nd par. to pg.5312 2nd par. [section titled: Products on-or on the way to-the market] Whole Document	67, 68, 70, 71, 84, 85 1-88
L	WO 2007/061613 A2 (GLYCOFI INC) 31 May 2007 Incorporated within D1 and D3	
L	HAMILTON, S.R. et al., 'Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins', 2006. SCIENCE, Vol.313, No.5792, pages 1441-1443. Incorporated within D2	
A	CUSKIN, F. et al., 'Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism', 2015. Nature, Vol.517, No.7533, pages 165-169. Whole Document	27, 76
L	DE POURCQ, K. et al., 'Engineering of glycosylation in yeast and other fungi: current state and perspectives', 2010. Appl Microbiol Biotechnol, Vol.87, No.5, pages 1617- 1631. Incorporated within D6	

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed.
 - b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).
☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT		International application No.	
Information on patent family members		PCT/US2022/038095	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
WO 2012/115903 A2	30 August 2012	WO 2012115903 A2	30 Aug 2012
		AU 2012220890 A1	29 Aug 2013
		CA 2827732 A1	30 Aug 2012
		CN 103764837 A	30 Apr 2014
		EP 2678440 A2	01 Jan 2014
		EP 2678440 B1	23 May 2018
		JP 2014518608 A	07 Aug 2014
		KR 20140114278 A	26 Sep 2014
		US 2013295608 A1	07 Nov 2013
		US 8936918 B2	20 Jan 2015
WO 2011/106389 A1	01 September 2011	None	
WO 2007/061613 A2	31 May 2007	None	
End of Annex			
Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.			