



(86) **Date de dépôt PCT/PCT Filing Date:** 2003/05/15  
(87) **Date publication PCT/PCT Publication Date:** 2003/11/27  
(45) **Date de délivrance/Issue Date:** 2014/10/14  
(85) **Entrée phase nationale/National Entry:** 2004/11/12  
(86) **N° demande PCT/PCT Application No.:** US 2003/015712  
(87) **N° publication PCT/PCT Publication No.:** 2003/097810  
(30) **Priorités/Priorities:** 2002/05/15 (US10/146,772);  
2002/09/09 (US10/241,742)

(51) **Cl.Int./Int.Cl. C12N 15/60** (2006.01),  
**C07H 21/00** (2006.01), **C12N 11/00** (2006.01),  
**C12N 15/55** (2006.01), **C12N 9/78** (2006.01),  
**C12N 9/88** (2006.01), **C12P 13/04** (2006.01),  
**C12P 7/02** (2006.01), **C12P 7/42** (2006.01),  
**C12P 7/56** (2006.01), **C12Q 1/34** (2006.01),  
**C12Q 1/527** (2006.01)

(72) **Inventeurs/Inventors:**  
DESANTIS, GRACE, US;  
SHORT, JAY M., US;  
BURK, MARK, US;  
WONG, KELVIN, US;  
FARWELL, ROBERT, US; ...

(54) **Titre : NITRILASES, ACIDES NUCLEIQUES LES CODANT ET LEURS PROCEDES DE PRODUCTION ET D'UTILISATION**  
(54) **Title: NITRILASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM**

(57) **Abrégé/Abstract:**

The invention relates to nitrilases and to nucleic acids encoding the nitrilases. In addition methods of designing new nitrilases and method of use thereof are also provided. The nitrilases have increased activity and stability at increased pH and temperature.



(72) **Inventeurs(suite)/Inventors(continued):** CHATMAN, KELLY, US

(73) **Propriétaires(suite)/Owners(continued):**VERENIUM CORPORATION, US

(74) **Agent:** NORTON ROSE FULBRIGHT CANADA LLP/S.E.N.C.R.L., S.R.L.



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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 November 2003 (27.11.2003)

PCT

(10) International Publication Number  
**WO 03/097810 A2**

(51) International Patent Classification<sup>7</sup>: **C12N**

(21) International Application Number: PCT/US03/15712

(22) International Filing Date: 15 May 2003 (15.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
10/146,772 15 May 2002 (15.05.2002) US  
10/241,742 9 September 2002 (09.09.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
US 10/241,742 (CIP)  
Filed on 9 September 2002 (09.09.2002)  
US 10/146,772 (CIP)  
Filed on 15 May 2002 (15.05.2002)

(71) Applicant (for all designated States except US): **DI-  
VERSA CORPORATION** [US/US]; 4955 Directors  
Place, San Diego, CA 92121 (US).

(72) Inventors; and  
(75) Inventors/Applicants (for US only): **DESANTIS, Grace**  
[CA/US]; 3520 Lebon Drive, Apt. 5223, San Diego,  
CA 92121 (US). **SHORT, Jay, M.** [US/US]; 681 Paseo  
Delicias, Box 7214, Rancho Santa Fe, CA 92067 (US).  
**BURK, Mark** [US/US]; 12634 Intermezzo Way, San  
Diego, CA 92130 (US). **WONG, Kevin** [US/US]; 11304

Trebol Street, San Diego, CA 92126 (US). **FARWELL,  
Robert** [US/US]; 4415 Terreno Court, San Diego, CA  
92124 (US). **CHATMAN, Kelly** [US/US]; P.O. Box 429,  
Solana Beach, CA 92075 (US).

(74) Agent: **EINHORN, Gregory, P.**; Fish & Richardson P.C.,  
4350 La Jolla Village Drive, Suite 500, San Diego, CA  
92122 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD,  
SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, YU, ZA, ZM, ZW.

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

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

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: NITRILASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM

(57) Abstract: The invention relates to nitrilases and to nucleic acids encoding the nitrilases. In addition methods of designing new nitrilases and method of use thereof are also provided. The nitrilases have increased activity and stability at increased pH and temperature.



**WO 03/097810 A2**



# **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_2\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

# **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_2\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.



# NITRILASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM

## FIELD OF THE INVENTION

[0003] The invention relates generally to the field of molecular biology, biochemistry and  
20 chemistry, and particularly to enzymatic proteins having nitrilase activity. The invention also  
relates to polynucleotides encoding the enzymes, and to uses of such polynucleotides and  
enzymes.

## BACKGROUND OF THE INVENTION

[0004] There are naturally occurring enzymes which have great potential for use in  
25 industrial chemical processes for the conversion of nitriles to a wide range of useful products  
and intermediates. Such enzymes include nitrilases which are capable of converting nitriles  
directly to carboxylic acids. Nitrilase enzymes are found in a wide range of mesophilic  
micro-organisms, including species of *Bacillus*, *Norcardia*, *Bacteridium*, *Rhodococcus*,



micro-organisms, including species of *Bacillus*, *Nocardia*, *Bacteridium*, *Rhodococcus*, *Micrococcus*, *Brevibacterium*, *Alcaligenes*, *Acinetobacter*, *Corynebacterium*, *Fusarium* and *Klebsiella*. Additionally, there are thermophilic nitrilases which exist in bacteria.

[0005] There are two major routes from a nitrile to an analogous acid: (1) a nitrilase catalyzes the direct hydrolysis of a nitrile to a carboxylic acid with the concomitant release of ammonia; or (2) a nitrile hydratase adds a molecule of water across the carbon-nitrogen bonding system to give the corresponding amide, which can then act as a substrate for an amidase enzyme which hydrolyzes the carbon-nitrogen bond to give the carboxylic acid product with the concomitant release of ammonia. The nitrilase enzyme therefore provides the more direct route to the acid.

[0006] A nitrile group offers many advantages in devising synthetic routes in that it is often easily introduced into a molecular structure and can be carried through many processes as a masked acid or amide group. This is only of use, however, if the nitrile can be unmasked at the relevant step in the synthesis. Cyanide represents a widely applicable C<sub>1</sub>-synthon (cyanide is one of the few water-stable carbanions) which can be employed for the synthesis of a carbon framework. However, further transformations of the nitrile thus obtained are impeded due to the harsh reaction conditions required for its hydrolysis using normal chemical synthesis procedures. The use of enzymes to catalyze the reactions of nitriles is attractive because nitrilase enzymes are able to effect reactions with fewer environmentally hazardous reagents and by-products than in many traditional chemical methods. Indeed, the chemoselective biocatalytic hydrolysis of nitriles represents a valuable alternative because it occurs at ambient temperature and near physiological pH.

[0007] The importance of asymmetric organic synthesis in drug design and discovery has fueled the search for new synthetic methods and chiral precursors which can be utilized in developing complex molecules of biological interest. One important class of chiral molecules is the  $\alpha$ -substituted carboxylic acids, which include the  $\alpha$ -amino acids. These molecules have long been recognized as important chiral precursors to a wide variety of complex biologically active molecules, and a great deal of research effort has been dedicated to the development of methods for the synthesis of enantiomerically pure  $\alpha$ -amino acids and chiral medicines.

[0008] Of particular use to synthetic chemists who make chiral medicines would be an enzyme system which is useful under non-sterile conditions, which is useful in non-biological laboratories, which is available in a form convenient for storage and use; which has broad



substrate specificity, which acts on poorly water soluble substrates; which has predictable product structure; which provides a choice of acid or amide product; and which is capable of chiral differentiation. Accordingly, there is a need for efficient, inexpensive, high-yield synthetic methods for producing enantiomerically pure  $\alpha$ -substituted carboxylic acids, such as, for example,  $\alpha$ -amino acids and  $\alpha$ -hydroxy acids.

[0009] In addition, often, the discovery or evolution of an enzyme to perform a particular transformation can be aided by the availability of a convenient high throughput screening or selection process. While a surrogate substrate may be used when an effective ultra high throughput (UHTP) screen is not available, it may be desirable to screen directly for an enzyme that performs specifically the desired transformation. The challenges of designing an UHTP screen is evident when, for example, the discovery or evolution program is aimed at uncovering a stereoselective transformation to generate only one stereoisomer or enantiomer. In this case, there is a paucity of high throughput screening methods available. While, the most straightforward method is to use chiral liquid or gas phase separation to separate the two enantiomers in question, often this approach does not afford the very high throughput capacity that is required. By using mass spectroscopy (MS) techniques, very high throughput screens are possible. However, when applied in a conventional manner, MS does not afford information on chirality or enantioselectivity.

[00010] Another approach is to chemically derivatize the enantiomeric mixture with a single enantiomer compound, thus generating a diasteriomer mixture of compounds that can be characterized by separation on an achiral stationary phase. Again, this is a cumbersome approach and does not lend itself well to high throughput screening.

[00011] Throughout this application, various publications are referenced by author and date. The disclosures of these publications

more fully describe the state of the art as known to those skilled therein as of the date of the invention described and claimed herein.

## SUMMARY OF THE INVENTION

[00012] The present invention is directed to an isolated or recombinant nucleic acid comprising nucleotides having a sequence at least about 50% identical to SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177,



179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, or variants thereof, wherein the nucleic acid encodes a polypeptide having a nitrilase activity. In alternative aspects of the invention, the nucleic acid comprises nucleotides having a sequence at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete identity (100% identical) to the SEQ ID NO: or variants thereof. Exemplary variants may include, for example, the following variations of SEQ ID NO:195, 205, 207, 209, OR 237, having one or more mutations: at positions 163-165 AAA, AAG, GGT, GGC, GGA, GGG, CAA, or CAG; at positions 178-180 GAA or GAG; at positions 331-333 TCT, TCC, TCA, TCG, AGT, or AGC; at positions 568-570 CAT, CAC, TCT, TCC, TCA, TCG, AGT, AGC, ACT, ACC, ACA, TCA, TAT, TAC, ATG or ACG; at positions 571-573 TTA, TTG, CTT, CTC, CTA, CTG, GTT, GTC, GTA, GTG, ATG, ACT, ACC, ACA, GAT, GAC, GGT, GGC, GGA, GGG, GAA, GAG, TAT, TAC, or ACG; at positions 595-597 GAA, GAG, TTA, TTG, CTT, CTC, CTA, or CTG; at positions 664-666 TTA, TTG, CTT, CTC, CTA, or CTG; or any combination thereof. In one aspect of the invention, the variants encode a polypeptide having improved or diminished enantioselectivity, for example, in the conversion of a 3-hydroxyglutarylnitrile (HGN) to (R)-4-Cyano-3-hydroxybutyrate, than the polypeptide encoded by the SEQ ID NO.

**[00013]** In one aspect of the invention, the nucleic acid comprises nucleotides having a sequence substantially identical to the SEQ ID NO: or variants thereof. In another aspect, the invention provides for an isolated or recombinant nucleic acid comprising consecutive nucleotides having a sequence at least 79 % identical to SEQ ID NO: 33, wherein the nucleic acid encodes a polypeptide having nitrilase activity. The invention provides for a fragment of the nucleic acid, wherein the fragment encodes a polypeptide having nitrilase activity. The invention also provides for an isolated or recombinant nucleic acid complementary to any of the nucleic acids. The invention also provides for an isolated or recombinant nucleic acid



that hybridizes to any one of the nucleic acids under stringent conditions. In one aspect, the stringent conditions comprise at least 50% formamide, and about 37°C to about 42°C.

[00014] The invention provides for a nucleic acid probe comprising from about 15 nucleotides to about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more nucleotides, wherein at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more consecutive nucleotides are at least 50% complementary to a nucleic acid target region within a nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, variants thereof, or their complements. In one aspect, the nucleic acid probe comprises consecutive nucleotides which are at least 55% complementary to the nucleic acid target region. In one aspect, the invention provides for a nucleic acid probe, wherein the consecutive nucleotides are at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more or 100% complementary to the nucleic acid target region. In another aspect, the nucleic acid consists essentially of from about 20 to about 50 nucleotides. In other aspects, the nucleic acid can be at least about 20, 25, 30, 35, 40, 45, 50, 75, 100, 150 nucleotides in length.

[00015] The invention provides for a nucleic acid vector capable of replication in a host cell, wherein the vector comprises the nucleic acid of the invention. The invention also provides for a host cell comprising the nucleic acid. The invention also provides for a host organism comprising the host cell. In one aspect, the host organism comprises a gram negative bacterium, a gram positive bacterium or a eukaryotic organism. In another aspect, the gram negative bacterium comprises *Escherichia coli*, or *Pseudomonas fluorescens*. In a



further aspect, the gram positive bacterium comprises *Streptomyces diversa*, *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, or *Bacillus subtilis*. In a further aspect, the eukaryotic organism comprises *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, or *Aspergillus niger*.

5 [00016] The invention provides for an isolated or recombinant nucleic acid encoding a polypeptide comprising amino acids having a sequence at least 50% identical to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 10 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 15 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, or variants thereof, wherein the polypeptide has nitrilase activity. In one aspect, the polypeptide comprises amino acids having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 20 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more or 100% identity to the SEQ ID NO: or variants thereof. Exemplary variants may include, for example, the following variations of SEQ ID NO:196, 206, 208, 210 or 238, having one or more mutations: at residue 55 lysine, glycine, or glutamine; at residue 60 glutamic acid; at residue 25 111 serine, at residue 190, serine, histidine, tyrosine or threonine; at residue 191, leucine, valine, methionine, aspartic acid, glycine, glutamic acid, tyrosine or threonine; at residue 199 glutamic acid or leucine; at residue 222 leucine; or any combination thereof.

[00017] The invention also provides for an isolated or recombinant nucleic acid encoding a polypeptide comprising at least 10 consecutive amino acids having a sequence identical to a 10 portion of an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154,



156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190,  
 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226,  
 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262,  
 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298,  
 5 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334,  
 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370,  
 372, 374, 376, 378, 380, 382, 384, 386, or variants thereof.

**[00018]** An isolated or recombinant polypeptide comprising amino acids having a  
 sequence at least about 50% identical to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24,  
 10 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74,  
 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116,  
 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152,  
 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188,  
 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224,  
 15 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260,  
 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296,  
 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332,  
 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368,  
 370, 372, 374, 376, 378, 380, 382, 384, 386, or variants thereof, wherein the polypeptide has  
 20 nitrilase activity. In one aspect of the invention, the polypeptide comprises amino acids  
 having a sequence at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%,  
 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%,  
 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%,  
 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more or 100% identical to the SEQ ID NO:  
 25 or variants thereof.

**[00019]** The invention provides an isolated or recombinant nucleic acid comprising  
 nucleotides having a sequence as set forth in any one of the following SEQ ID NOS:1, 3, 5, 7,  
 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59,  
 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105,  
 30 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141,  
 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177,  
 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249,



251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, and variants thereof

(hereinafter referred to as "Group A nucleic acids"). The invention is also directed to nucleic acids having specified minimum percentages of sequence identity to any of the Group A nucleic acids sequences.

**[00020]** In another aspect, the invention provides an isolated (purified) or recombinant polypeptide comprising amino acid residues having a sequence as set forth in any one of the

following SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, and variants thereof, (hereinafter referred to as "Group B amino acid

sequences"). The invention is also directed to purified polypeptides having specified minimum percentages of sequence identity to any of the Group B amino acid sequences.

**[00021]** The invention provides for a fragment of the polypeptide which is at least 50 amino acids in length, and wherein the fragment has nitrilase activity. Furthermore, the

invention provides for a peptidomimetic of the polypeptide or a fragment thereof having nitrilase activity. The invention provides for a codon-optimized polypeptide or a fragment thereof, having nitrilase activity, wherein the codon usage is optimized for a particular organism or cell. Narum et al. *Infect. Immun.* 2001 Dec, 69(12):7250-3 describes codon-optimization in the mouse system. Outchkourov et al. *Protein Expr. Purif.* 2002 Feb;

24(1):18-24 describes codon-optimization in the yeast system. Feng et al. *Biochemistry* 2000 Dec 19, 39(50):15399-409 describes codon-optimization in *E. coli*. Humphreys et al. *Protein Expr. Purif.* 2000 Nov, 20(2):252-64 describes how codon usage affects secretion in *E. coli*.



[00022] In one aspect, the organism or cell comprises a gram negative bacterium, a gram positive bacterium or a eukaryotic organism. In another aspect of the invention, the gram negative bacterium comprises *Escherichia coli*, or *Pseudomonas fluorescens*. In another aspect of the invention, the gram positive bacterium comprise *Streptomyces diversa*,

5 *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, or *Bacillus subtilis*. In another aspect of the invention, the eukaryotic organism comprises *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, or *Aspergillus niger*.

[00023] In another aspect, the invention provides for a purified antibody that specifically  
10 binds to the polypeptide of the invention or a fragment thereof, having nitrilase activity. In one aspect, the invention provides for a fragment of the antibody that specifically binds to a polypeptide having nitrilase activity.

[00024] The invention provides for an enzyme preparation which comprises at least one of the polypeptides of the invention, wherein the preparation is liquid or dry. The enzyme  
15 preparation includes a buffer, cofactor, or second or additional protein. In one aspect the preparation is affixed to a solid support. In one aspect of the invention, the solid support can be a gel, a resin, a polymer, a ceramic, a glass, a microelectrode and any combination thereof. In another aspect, the preparation can be encapsulated in a gel or a bead.

[00025] The invention further provides for a composition which comprises at least one  
20 nucleic acid of the invention which comprises at least one polypeptide of the invention or a fragment thereof, or a peptidomimetic thereof, having nitrilase activity, or any combination thereof.

[00026] The invention provides for a method for hydrolyzing a nitrile to a carboxylic acid comprising contacting the molecule with at least one polypeptide of the invention or a  
25 fragment thereof, or a peptidomimetic thereof, having nitrilase activity, under conditions suitable for nitrilase activity. In one aspect, the conditions comprise aqueous conditions. In another aspect, the conditions comprise a pH of about 8.0 and/or a temperature from about 37° C to about 45° C.

[00027] The invention provides for a method for hydrolyzing a cyanohydrin moiety or an  
30 aminonitrile moiety of a molecule, the method comprising contacting the molecule with at least one polypeptide of the invention, or a fragment thereof, or a peptidomimetic thereof, having nitrilase activity, under conditions suitable for nitrilase activity.



[00028] The invention provides for a method for making a chiral  $\alpha$ -hydroxy acid molecule, a chiral amino acid molecule, a chiral  $\beta$ -hydroxy acid molecule, or a chiral gamma-hydroxy acid molecule, the method comprising admixing a molecule having a cyanohydrin moiety or an aminonitrile moiety with at least one polypeptide having an amino acid sequence at least 50% identical to any one of the Group B amino acid sequences or a fragment thereof, or a peptidomimetic thereof, having enantio-selective nitrilase activity. In one aspect, the chiral molecule is an (*R*)-enantiomer. In another aspect, the chiral molecule is an (*S*)-enantiomer. In one aspect of the invention, one particular enzyme can have R-specificity for one particular substrate and the same enzyme can have S-specificity for a different particular substrate.

[00029] The invention also provides for a method for making a composition or an intermediate thereof, the method comprising admixing a precursor of the composition or intermediate, wherein the precursor comprises a cyanohydrin moiety or an aminonitrile moiety, with at least one polypeptide of the invention or a fragment or peptidomimetic thereof having nitrilase activity, hydrolyzing the cyanohydrin or the aminonitrile moiety in the precursor thereby making the composition or the intermediate thereof. In one aspect, the composition or intermediate thereof comprises (*S*)-2-amino-4-phenyl butanoic acid. In a further aspect, the composition or intermediate thereof comprises an L-amino acid. In a further aspect, the composition comprises a food additive or a pharmaceutical drug.

[00030] The invention provides for a method for making an (*R*)-ethyl 4-cyano-3-hydroxybutyric acid, the method comprising contacting a hydroxyglutaryl nitrile with at least one polypeptide having an amino acid sequence of the Group B amino acid sequences, or a fragment or peptidomimetic thereof having nitrilase activity that selectively produces an (*R*)-enantiomer, so as to make (*R*)-ethyl 4-cyano-3-hydroxybutyric acid. In one aspect, the *ee* is at least 95% or at least 99%. In another aspect, the hydroxyglutaryl nitrile comprises 1,3-dicyano-2-hydroxy-propane or 3-hydroxyglutaronitrile. In a further aspect, the polypeptide has an amino acid sequence of any one of the Group B amino acid sequences, or a fragment or peptidomimetic thereof having nitrilase activity.

[00031] The invention also provides a method for making an (*S*)-ethyl 4-cyano-3-hydroxybutyric acid, the method comprising contacting a hydroxyglutaryl nitrile with at least one polypeptide having an amino acid sequence of the Group B amino acid sequences, or a fragment or peptidomimetic thereof having nitrilase activity that selectively produces an (*S*)-enantiomer, so as to make (*S*)-ethyl 4-cyano-3-hydroxybutyric acid.



[00032] The invention provides a method for making an (*R*)-mandelic acid, the method comprising admixing a mandelonitrile with at least one polypeptide having an amino acid sequence of any one of the Group B amino acid sequences or any fragment or peptidomimetic thereof having appropriate nitrilase activity. In one aspect, the (*R*)-mandelic acid comprises (*R*)-2-chloromandelic acid. In another aspect, the (*R*)-mandelic acid comprises an aromatic ring substitution in the *ortho*-, *meta*-, or *para*- positions; a 1-naphthyl derivative of (*R*)-mandelic acid, a pyridyl derivative of (*R*)-mandelic acid or a thienyl derivative of (*R*)-mandelic acid or a combination thereof.

[00033] The invention provides a method for making an (*S*)-mandelic acid, the method comprising admixing a mandelonitrile with at least one polypeptide having an amino acid sequence of Group B sequences or any fragment or peptidomimetic thereof having nitrilase activity. In one aspect, the (*S*)-mandelic acid comprises (*S*)-methyl benzyl cyanide and the mandelonitrile comprises (*S*)-methoxy-benzyl cyanide. In one aspect, the (*S*)-mandelic acid comprises an aromatic ring substitution in the *ortho*-, *meta*-, or *para*- positions; a 1-naphthyl derivative of (*S*)-mandelic acid, a pyridyl derivative of (*S*)-mandelic acid or a thienyl derivative of (*S*)-mandelic acid or a combination thereof.

[00034] The invention also provides a method for making an (*S*)-phenyl lactic acid derivative or an (*R*)-phenyllactic acid derivative, the method comprising admixing a phenyllactonitrile with at least one polypeptide selected from the group of the Group B amino acid sequences or any active fragment or peptidomimetic thereof that selectively produces an (*S*)-enantiomer or an (*R*)-enantiomer, thereby producing an (*S*)-phenyl lactic acid derivative or an (*R*)-phenyl lactic acid derivative.

[00035] The invention provides for a method for making the polypeptide of the invention or a fragment thereof, the method comprising (a) introducing a nucleic acid encoding the polypeptide into a host cell under conditions that permit production of the polypeptide by the host cell, and (b) recovering the polypeptide so produced.

[00036] The invention provides for a method for generating a nucleic acid variant encoding a polypeptide having nitrilase activity, wherein the variant has an altered biological activity from that which naturally occurs, the method comprising (a) modifying the nucleic acid by (i) substituting one or more nucleotides for a different nucleotide, wherein the nucleotide comprises a natural or non-natural nucleotide; (ii) deleting one or more nucleotides, (iii) adding one or more nucleotides, or (iv) any combination thereof. In one aspect, the non-natural nucleotide comprises inosine. In another aspect, the method further



comprises assaying the polypeptides encoded by the modified nucleic acids for altered nitrilase activity, thereby identifying the modified nucleic acid(s) encoding a polypeptide having altered nitrilase activity. In one aspect, the modifications of step (a) are made by PCR, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis, ligase chain reaction, *in vitro* mutagenesis, ligase chain reaction, oligonucleotide synthesis, any DNA-generating technique and any combination thereof. In another aspect, the method further comprises at least one repetition of the modification step (a).

[00037] The invention further provides a method for making a polynucleotide from two or more nucleic acids, the method comprising: (a) identifying regions of identity and regions of diversity between two or more nucleic acids, wherein at least one of the nucleic acids comprises a nucleic acid of the invention; (b) providing a set of oligonucleotides which correspond in sequence to at least two of the two or more nucleic acids; and, (c) extending the oligonucleotides with a polymerase, thereby making the polynucleotide.

[00038] The invention further provides a screening assay for identifying a nitrilase, the assay comprising: (a) providing a plurality of nucleic acids or polypeptides comprising at least one of the nucleic acids of the invention, or at least one of the polypeptides of the invention; (b) obtaining polypeptide candidates to be tested for nitrilase activity from the plurality; (c) testing the candidates for nitrilase activity; and (d) identifying those polypeptide candidates which are nitrilases. In one aspect, the method further comprises modifying at least one of the nucleic acids or polypeptides prior to testing the candidates for nitrilase activity. In another aspect, the testing of step (c) further comprises testing for improved expression of the polypeptide in a host cell or host organism. In a further aspect, the testing of step (c) further comprises testing for nitrilase activity within a pH range from about pH 3 to about pH 12. In a further aspect, the testing of step (c) further comprises testing for nitrilase activity within a pH range from about pH 5 to about pH 10. In another aspect, the testing of step (c) further comprises testing for nitrilase activity within a temperature range from about 4 °C to about 80 °C. In another aspect, the testing of step (c) further comprises testing for nitrilase activity within a temperature range from about 4 °C to about 55 °C. In another aspect, the testing of step (c) further comprises testing for nitrilase activity which results in an enantioselective



reaction product. In another aspect, the testing of step (c ) further testing for nitrilase activity which results in a regio-selective reaction product.

[00039] The invention provides for use of the nucleic acids of the invention, or a fragment or peptidomimetic thereof having nitrilase activity, in a process designed to optimize one  
5 aspect of the gene or one aspect of the polypeptide encoded by the gene. In one aspect, the process comprises introducing modifications into the nucleotide sequence of the nucleic acid. In another aspect, the modifications are introduced by PCR, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble  
0 mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis, ligase chain reaction, *in vitro* mutagenesis, ligase chain reaction, oligonucleotide synthesis, any other DNA-generating technique or any combination thereof. In a further aspect, the process is repeated.

[00040] The invention provides for use of the polypeptide of the invention, or a fragment  
5 or peptidomimetic thereof having nitrilase activity, in an industrial process. In one aspect, the process is for production of a pharmaceutical composition, the process is for production of a chemical, the process is for production of a food additive, the process is catalyzing the breakdown of waste, or the process is production of a drug intermediate. In a further aspect, the process comprises use of the polypeptide to hydrolyze a hydroxyglutaryl nitrile substrate.  
0 In a further aspect, the process is for production of LIPITOR™. In another aspect, the polypeptide used comprises a polypeptide having consecutive amino acids of the sequence SEQ ID NO:44, 196, 208, 210, or 238 or a fragment thereof having nitrilase activity. In another aspect, the process is production of a detergent. In another aspect, the process is production of a food product.

5 [00041] The invention provides for use of a nucleic acid of the invention, or a fragment thereof encoding a polypeptide having nitrilase activity, in the preparation of a transgenic organism.

[00042] The invention provides for a kit comprising (a) the nucleic acid of the inventions, or a fragment thereof encoding a polypeptide having nitrilase activity, or (b) the polypeptide  
0 of the invention, or a fragment or a peptidomimetic thereof having nitrilase activity, or a combination thereof; and (c) a buffer.

[00043] The invention provides for a method for modifying a molecule comprising: (a) mixing a polypeptide of the invention or a fragment or peptidomimetic thereof having



nitrilase activity, with a starting molecule to produce a reaction mixture; (b) reacting the starting molecule with the polypeptide to produce the modified molecule.

[00044] The invention provides for a method for identifying a modified compound comprising: (a) admixing a polypeptide of the invention, or a fragment or peptidomimetic thereof having nitrilase activity, with a starting compound to produce a reaction mixture and thereafter a library of modified starting compounds; (b) testing the library to determine whether a modified starting compound is present within the library which exhibits a desired activity; (c) identifying the modified compound exhibiting the desired activity.

[00045] The invention provides a screening assay for enantioselective transformation comprising: (a) providing a molecule having two prochiral or enantiotopic moieties; (b) labeling at least one prochiral or enantiotopic moiety of the molecule; (b) modifying at least one of the two moieties by a selective catalyst; and (c) detecting results by mass spectroscopy. The screening assay can be used to determine or monitor the % enantiomeric excess (ee) or determine the % diastereomeric excess (de). An exemplary label useful in the assay is a heavier isotope or a lighter isotope. The selective catalyst useful in the assay can be an enzyme. The screening assay may be performed with both moieties labeled. The screening assay may be performed in both directions, *i.e.*, from the reactants to the products as well as from the products to the reactants.

[00046] The invention provides for a computer readable medium having stored thereon a nucleic acid of the invention, e.g., a nucleic acid comprising at least one nucleotide sequence selected from the group consisting of: SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, and variants thereof, and/or at least one amino acid sequence selected from the group consisting of: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70,



72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, and variants thereof.

10 **[00047]** The invention provides for a computer system comprising a processor and a data storage device, wherein the data storage device has stored thereon a nucleic acid of the invention, e.g., a nucleic acid comprising at least one nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 15 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 20 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, and variants thereof, and/or at least one amino acid sequence selected from the group consisting of: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 25 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 30 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374,



376, 378, 380, 382, 384, 386, and variants thereof. In one aspect, the computer system further comprises a sequence comparison algorithm and a data storage device having at least one reference sequence stored thereon. In another aspect, the sequence comparison algorithm comprises a computer program that identifies polymorphisms.

5 [00048] The invention provides for a method for identifying a feature in a sequence which comprises: (a) inputting the sequence into a computer; (b) running a sequence feature identification program on the computer so as to identify a feature within the sequence; and (c) identifying the feature in the sequence, wherein the sequence comprises a nucleic acid of the invention, e.g., a nucleic acid comprising at least one of SEQ ID NOS:1-386, its variants, or  
10 any combination thereof.

[00049] The invention provides for an assay for identifying a functional fragment of a polypeptide which comprises: (a) obtaining a fragment of at least one polypeptide of the invention; (b) contacting at least one fragment from step (a) with a substrate having a cyanohydrin moiety or an aminonitrile moiety under reaction conditions suitable for nitrilase  
15 activity; (c) measuring the amount of reaction product produced by each at least one fragment from step (b); and (d) identifying the at least one fragment which is capable of producing a nitrilase reaction product; thereby identifying a functional fragment of the polypeptide. In one aspect, the fragment of step (a) is obtained by synthesizing the fragment. In another aspect, the fragment of step (a) is obtained by fragmenting the polypeptides.

20 [00050] The invention provides for an assay for identifying a functional variant of a polypeptide which comprises: (a) obtaining at least one variant of at least one polypeptide of the invention; (b) contacting at least one variant from step (a) with a substrate having a cyanohydrin moiety or an aminonitrile moiety under reaction conditions suitable for nitrilase activity; (c) measuring the amount of reaction product produced by each at least one variant  
25 from step (b); and (d) identifying the at least one variant which is capable of producing a nitrilase reaction product; thereby identifying a functional variant of the polypeptide.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[00051] Figure 1 shows chemical reaction schemes wherein stereoselective nitrilases hydrolyze a cyanohydrin or an aminonitrile to produce a chiral  $\alpha$ -hydroxy acid or  $\alpha$ -amino  
30 acid.

[00052] Figure 2 illustrates an OPA based cyanide detection assay used for identifying the presence of nitrilase activity.



[00053] Figure 3 is an illustration of a spectroscopic system for the detection and quantification of  $\alpha$ -hydroxy acids based on stereoselective lactate dehydrogenases.

[00054] Figure 4 is an illustration of a spectroscopic system for the detection and quantification of  $\alpha$ -amino acids based on stereoselective amino acid oxidase.

5 [00055] Figure 5 is a flow diagram illustrating the steps of a nitrilase screening method.

[00056] Figures 6A-6E are chromatograms characteristic of the substrate and product combination for D-phenylglycine showing a blank sample (Fig. 6A), an enzymatic reaction sample (Fig. 6B); a negative control consisting of cell lysate in buffer (Fig. 6C); a chiral analysis of phenylglycine (Fig. 6D); and coelution of the nitrile peak with the D-enantiomer  
10 (Fig. 6E).

[00057] Figures 7A-7E illustrate chromatograms which are characteristic of substrate and product combinations for (*R*)-2-chloromandelic acid. Fig. 7A shows only 2-chloromandelonitrile in buffer; Fig. 7B shows a chloromandelic acid standard. The chromatogram in Fig. 7C shows the appearance of product and the reduction of substrate  
15 peaks.

[00058] Figures 8A-8B illustrate chromatograms characteristic of substrate and product combinations for (*S*)-phenyllactic acid.

[00059] Figures 9A-9B illustrate chromatograms characteristic of substrate and product combinations for L-2-methylphenylglycine.

20 [00060] Figures 10A-10C illustrate chromatograms characteristic of substrate and product combinations for L-*tert*-leucine.

[00061] Figures 11A-11C illustrate chromatograms characteristic of substrate and product combinations for (*S*)-2-amino-6-hydroxy hexanoic acid.

[00062] Figures 12A-12D illustrate chromatograms characteristic of substrate and product  
25 combinations for 4-methyl-D-leucine and 4-methyl-L-leucine.

[00063] Figures 13A-13B illustrate chromatograms characteristic of substrate and product combinations for (*S*)-cyclohexylmandelic acid.

[00064] Figures 14A-14B illustrate two exemplary standard curves for quantitation in connection with the screening assay of the invention.

30 [00065] Figure 15 illustrates selected compounds that can be produced from a nitrilase-catalyzed reaction using an enzyme and/or a method of the invention.

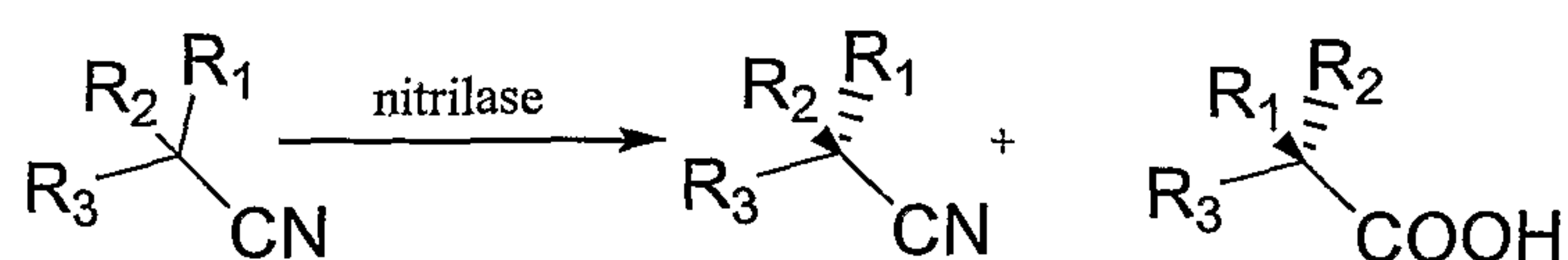
[00066] Figure 16 illustrates selected compounds that can be produced from a nitrilase-catalyzed reaction using an enzyme and/or a method of the invention.



[00067] Figure 17 illustrates an exemplary nitrilase reaction of the invention.

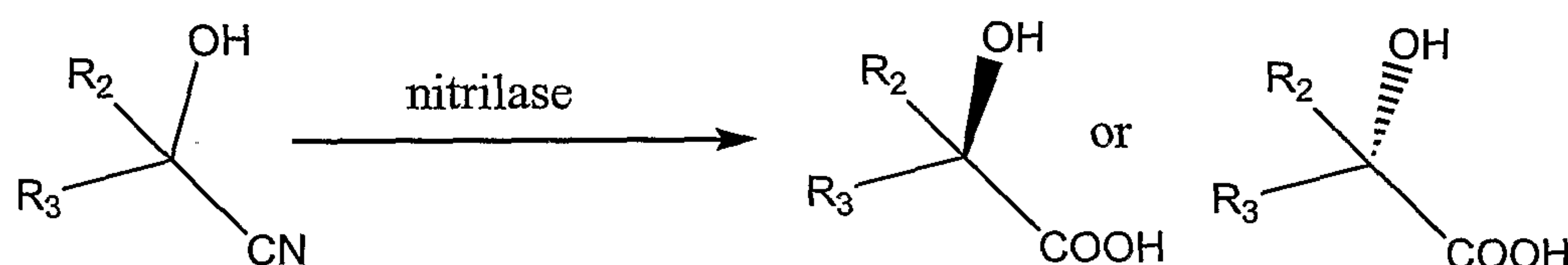
### DETAILED DESCRIPTION OF THE INVENTION

[00068] The present invention relates to nitrilases, nucleic acids encoding nitrilases, and  
 5 uses therefor. As used herein, the term “nitrilase” encompasses any polypeptide having any nitrilase activity, for example, the ability to hydrolyze nitriles into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantioselective aromatic and aliphatic amino acids or hydroxy acids.

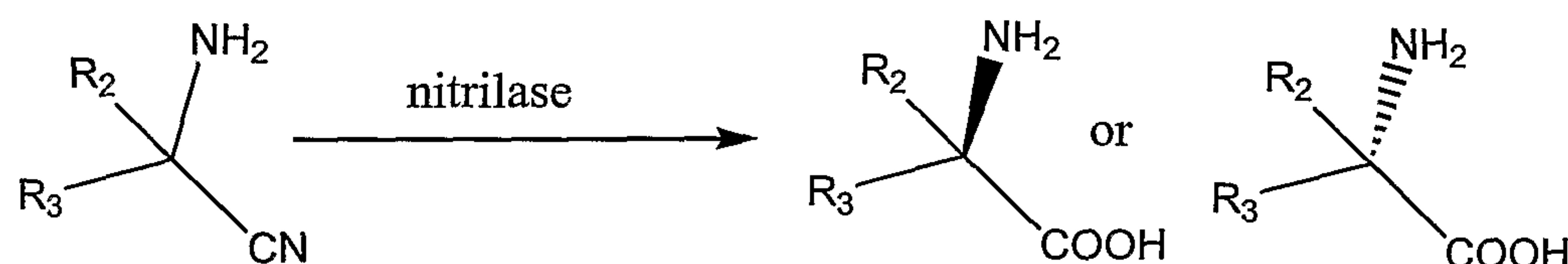


[00069] Nitrilase chemistry is as follows:

10 [00070] A nitrilase reaction for the preparation of hydroxy acids is as follows:



[00071] A nitrilase reaction for the preparation of amino acids is as follows:



15 [00072] In addition, in each of the foregoing hydrolysis reactions, two water molecules are consumed and one ammonia molecule is released.

[00073] There are several different types of assays which can be performed to test for the presence of nitrilase activity in a sample or to test whether a particular polypeptide exhibits nitrilase activity. For example, assays can detect the presence or absence of products or by-  
 20 products from a chemical reaction catalyzed by a nitrilase. For example, the presence of nitrilase activity can be detected by the production of α-hydroxy acids or α-amino acids from, respectively, cyanohydrins or aminonitriles, and the level of nitrilase activity can be



quantified by measuring the relative quantities of the reaction products produced. Figure 1 shows chemical reaction schemes using stereoselective nitrilases to create chiral  $\alpha$ -hydroxy acids or  $\alpha$ -amino acids in high yield. The starting material is an aldehyde or an imine which is produced from an aldehyde by reaction with ammonia. Reaction of the aldehyde or imine with hydrogen cyanide results in the production of enantiomeric mixtures of the corresponding cyanohydrins and aminonitriles. A stereoselective nitrilase can then be used to stereoselectively convert one enantiomer into the corresponding  $\alpha$ -hydroxy acid or  $\alpha$ -amino acid. Figure 3 illustrates schematically the stereoselective nitrilase-dependent production and spectrophotometric detection of  $\alpha$ -hydroxy acids based on lactate dehydrogenase conversion of the  $\alpha$ -hydroxy acids to the corresponding  $\alpha$ -keto acids and concomitant oxidation-reduction of a detectable dye. Figure 4 illustrates schematically the stereoselective nitrilase-dependent production and spectrophotometric detection of  $\alpha$ -amino acids based on amino acid oxidase conversion of the  $\alpha$ -amino acids to the corresponding  $\alpha$ -keto acids and concomitant oxidation-reduction of a detectable dye.

[00074] Nitrilases contemplated for use in the practice of the present invention include those which stereoselectively hydrolyze nitriles or cyanohydrins into their corresponding acids and ammonia. In one aspect, nitrilases of the invention can stereoselectively hydrolyze nitriles or cyanohydrins into their corresponding acids and ammonia. Nitrilases include, for example, nitrilases of the invention, e.g., those set forth in the Group B amino acid sequences. Some nitrilases which stereoselectively hydrolyze their substrates are set forth in the Tables hereinbelow.

[00075] The nitrilases of the invention share the following additional characteristics:

- (1) full-length amino acid sequences from about 333 amino acids to about 366 amino acids,
- (2) aggregation and activity as homo-multimers of about 2 subunits to about 16 subunits,
- (3) presence of a catalytic triad of the consecutive amino acids Glu-Lys-Cys, (4) pH optima from about pH 5 to about pH 9, and (5) temperature optima from about 0°C to about 100°C, or from about 40°C to about 50°C.

#### Consensus Sequences Among New Nitrilases

[00076] The nitrilases disclosed herein were studied using bioinformatics and sequence comparison programs and the following consensus information was collected. Three regions of conserved motifs were identified within the nitrilase polypeptides. These correspond to the catalytic triad (E-K-C) present in nitrilase enzymes. (H. Pace and C. Brenner (Jan. 15,



2001) "The Nitrilase Superfamily: classification, structure and function" *Genome Biology* Vol. 2, No. 1, pp 1-9.)

[00077] The abbreviations used herein are conventional one letter codes for the amino acids: A, alanine; B, asparagine or aspartic acid; C, cysteine; D aspartic acid; E, glutamate, glutamic acid; F, phenylalanine; G, glycine; H histidine; I isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; Z, glutamine or glutamic acid. See L. Stryer, *Biochemistry*, 1988, W. H. Freeman and Company, New York.

[00078] The computer sequence comparisons made among the nitrilase polypeptide sequences of the invention resulted in the identification of these motifs within each amino acid sequence:

|   |   |          |   |          |   |          |          |   |   |   |   |   |   |          |   |          |   |   |   |   |
|---|---|----------|---|----------|---|----------|----------|---|---|---|---|---|---|----------|---|----------|---|---|---|---|
| F | P | <u>E</u> | t | <i>f</i> | r | <u>R</u> | <u>K</u> | L | . | P | T | L | . | <u>C</u> | W | <u>E</u> | h | . | . | P |
|---|---|----------|---|----------|---|----------|----------|---|---|---|---|---|---|----------|---|----------|---|---|---|---|

[00079] The following residues (those that are underlined) are completely conserved among all of the identified nitrilases: the third amino acid in the first motif or region (E, glutamate); the second residue in the second motif (R, arginine); the third residue in the second motif (K, lysine); the third residue in the third motif (C, cysteine); and the fifth residue in the third motif (E, glutamate).

[00080] In the boxes, upper case letters indicate 90% or greater consensus among the nitrilases of the invention, while lower case letters indicate 50% or greater consensus. An italicized letter indicates 30% or greater consensus among the nitrilases of the invention. A dot in a box indicates a residue which is not conserved.

[00081] The sequences of nitrilases in the nitrilase branch of the nitrilase superfamily were described as having a catalytic triad in the Pace and Brenner article (*Genome Biology*, 2001, Vol. 2, No. 1, pp. 1-9). However, the catalytic triad regions of the nitrilases of this invention differ from those previously identified in the Pace and Brenner reference in the following ways:

[00082] Differences in the first motif: The F in the first box of the first motif is conserved in 90% of the nitrilases of the invention, rather than in only 50% of those previously identified. The fourth residue of the first motif is a "t", threonine in the nitrilases of this invention, and it is found at 50% or greater consensus. However, that residue was identified by Pace and Brenner as "a" (alanine). The last residue of the first motif was



identified as "f" (phenylalanine) and was indicated to occur at 50% or greater consensus. However, the nitrilases of this invention only show "f" (phenylalanine occurring at 30% consensus.

[00083] Differences in the second motif: There is an "r" (arginine) in the first box of

5 the second motif of the nitrilases of this invention. However, the Pace and Brenner consensus shows an "h" (histidine) in that position. The "R" (arginine) in the second box is completely conserved in the nitrilases of the present invention, however that residue only appears at 90% consensus in the Pace and Brenner reference. The "L" (leucine) in the fourth box of the second motif is conserved in 90% or more of the nitrilases of this invention.

10 However, the Pace and Brenner nitrilases only showed conservation of that residue in 50% of the sequences. Similarly, the "P" (proline) at the sixth box of the second motif is conserved in 90% or more of the nitrilases of this invention. However, the Pace and Brenner nitrilases only showed conservation of that residue in 50% of the sequences.

[00084] Differences in the third motif: The "L" in the first box is conserved at 90% or

15 greater in the nitrilases of the invention. However, the Pace and Brenner reference only shows that residue appearing 50% of the time. Finally, the sixth box in the third motif in the nitrilases of the invention show a histidine 50% of the time or more. However, the Pace and Brenner reference indicates that that position shows an asparagine ("n") 50% of the time.

[00085] The invention provides for an isolated polypeptide having nitrilase activity which

20 polypeptide comprises three regions, wherein the first region comprises five amino acids and wherein the first amino acid of the first region is F and the fourth amino acid of the first region is T. The invention also provides for an isolated polypeptide having nitrilase activity which polypeptide comprises three regions, wherein the second region comprises seven amino acids and wherein the first amino acid of the second region is R, wherein the second amino acid of the second region is R, and wherein the sixth amino acid of the second region is P. The invention also provides for an isolated polypeptide having nitrilase activity which polypeptide comprises three regions, wherein the third region comprises nine amino acids and wherein the first amino acid of the third region is L and the sixth amino acid of the third region is H.

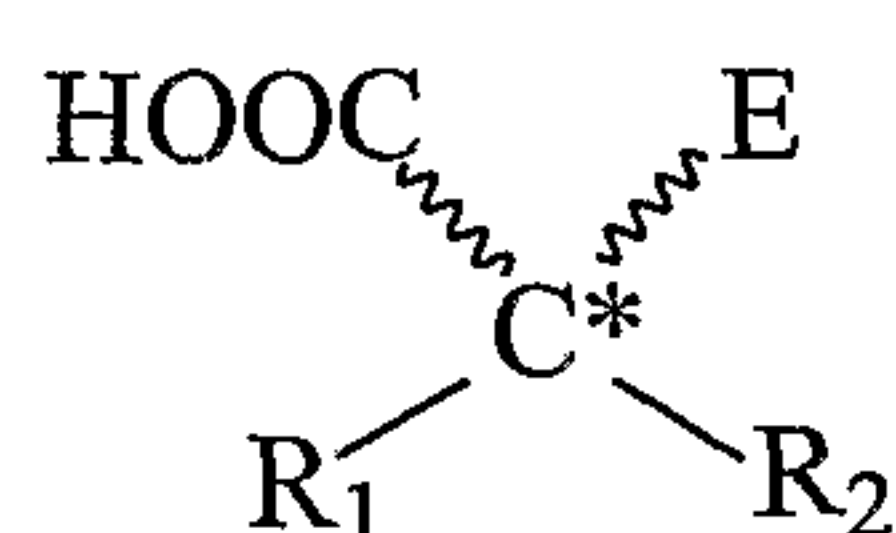
30 [00086] The invention also provides for an isolated polypeptide having nitrilase activity which polypeptide comprises three consensus subsequences, wherein the first consensus subsequence is FPETF, wherein the second consensus subsequence is RRKLXPT, and wherein the third consensus subsequence is LXCWEHXXP.



[00087] The invention also provides for an isolated polypeptide having nitrilase activity which polypeptide comprises three consensus subsequences, wherein the first consensus subsequence is FPEXX, wherein the second consensus subsequence is XRKLXPT, and wherein the third consensus subsequence is LXCWEXXXP.

5

[00088] In accordance with the present invention, methods are provided for producing enantiomerically pure  $\alpha$ -substituted carboxylic acids. The enantiomerically pure  $\alpha$ -substituted carboxylic acids produced by the methods of the present invention have the following structure:



10

wherein:

$\text{R}_1 \neq \text{R}_2$  and  $\text{R}_1$  and  $\text{R}_2$  are otherwise independently -H, substituted or unsubstituted alkyl, alkenyl, alkynyl, aryl, heteroaryl, cycloalkyl, or heterocyclic, wherein said substituents are lower alkyl, hydroxy, alkoxy, amino, mercapto, cycloalkyl, heterocyclic, aryl, heteroaryl, aryloxy, or halogen or optionally  $\text{R}_1$  and  $\text{R}_2$  are directly or indirectly covalently joined to form a functional cyclic moiety, and E is  $-\text{N}(\text{R}_x)_2$  or  $-\text{OH}$ , wherein each  $\text{R}_x$  is independently -H or lower alkyl.

15

[00089] As used herein, the term "alkyl" refers to straight or branched chain or cyclic hydrocarbon groups of from 1 to 24 carbon atoms, including methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, tert-butyl, n-pentyl, n-hexyl, and the like. The term "lower alkyl" refers to monovalent straight or branched chain or cyclic radicals of from one to about six carbon atoms.

20

[00090] As used herein, "alkenyl" refers to straight or branched chain or cyclic hydrocarbon groups having one or more carbon-carbon double bonds, and having in the range of about 2 to about 24 carbon atoms.

25

[00091] As used herein, "alkynyl" refers to straight or branched chain or cyclic hydrocarbon groups having at least one carbon-carbon triple bond, and having in the range of about 2 to about 24 carbon atoms.

[00092] As used herein, "cycloalkyl" refers to cyclic hydrocarbon groups containing in the range of about 3 to about 14 carbon atoms.

30



[00093] As used herein, "heterocyclic" refers to cyclic groups containing one or more heteroatoms (*e.g.*, N, O, S, P, Se, B, etc.) as part of the ring structure, and having in the range of about 3 to about 14 carbon atoms.

[00094] As used herein, "aryl" refers to aromatic groups (*i.e.*, cyclic groups with conjugated double-bond systems) having in the range of about 6 to about 14 carbon atoms.

[00095] As used herein with respect to a chemical group or moiety, the term "substituted" refers to such a group or moiety further bearing one or more non-hydrogen substituents.

Examples of such substituents include, without limitation, oxy (*e.g.*, in a ketone, aldehyde, ether, or ester), hydroxy, alkoxy (of a lower alkyl group), amino, thio, mercapto (of a lower alkyl group), cycloalkyl, substituted cycloalkyl, heterocyclic, substituted heterocyclic, aryl, substituted aryl, heteroaryl, substituted heteroaryl, aryloxy, substituted aryloxy, halogen, trifluoromethyl, cyano, nitro, nitrene, amino, amido, -C(O)H, acyl, oxyacyl, carboxyl, carbamate, sulfonyl, sulfonamide, sulfonyl, and the like.

[00096] In preferred aspects, the enantiomerically pure  $\alpha$ -substituted carboxylic acid produced by the methods of the present invention is an  $\alpha$ -amino acid or  $\alpha$ -hydroxy acid. In some aspects, the enantiomerically pure  $\alpha$ -amino acid is D-phenylalanine, D-phenylglycine, L-methylphenylglycine, L-*tert*-leucine, D-alanine, or D-hydroxynorleucine ((S)-2-amino-6-hydroxy hexanoic acid), R-pantolactone, 2-chloromandelic acid, or (S)- or (R)-mandelic acid and the enantiomerically pure  $\alpha$ -hydroxy acid is (S)-cyclohexylmandelic acid. As used herein, a "small molecule" encompasses any molecule having a molecular weight from at least 25 Daltons.

[00097] The term "about" is used herein to mean approximately, roughly, around, or in the region of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent up or down (higher or lower).

[00098] As used herein, the word "or" means any one member of a particular list and also includes any combination of members of that list.

[00099] The phrase "nucleic acid" as used herein refers to a naturally occurring or synthetic oligonucleotide or polynucleotide, whether DNA or RNA or DNA-RNA hybrid, single-stranded or double-stranded, sense or antisense, which is capable of hybridization to a complementary nucleic acid by Watson-Crick base-pairing. Nucleic acids of the invention can also include nucleotide analogs (*e.g.*, BrdU), and non-phosphodiester internucleoside



linkages (*e.g.*, peptide nucleic acid (PNA) or thiodiester linkages). In particular, nucleic acids can include, without limitation, DNA, RNA, cDNA, gDNA, ssDNA or dsDNA or any combination thereof. In some aspects, a "nucleic acid" of the invention includes, for example, a nucleic acid encoding a polypeptide as set forth in the Group B amino acid sequences, and variants thereof. The phrase "a nucleic acid sequence" as used herein refers to a consecutive list of abbreviations, letters, characters or words, which represent nucleotides. In one aspect, a nucleic acid can be a "probe" which is a relatively short nucleic acid, usually less than 100 nucleotides in length. Often a nucleic acid probe is from about 50 nucleotides in length to about 10 nucleotides in length. A "target region" of a nucleic acid is a portion of a nucleic acid that is identified to be of interest.

[000100] A "coding region" of a nucleic acid is the portion of the nucleic acid which is transcribed and translated in a sequence-specific manner to produce into a particular polypeptide or protein when placed under the control of appropriate regulatory sequences. The coding region is said to encode such a polypeptide or protein.

[000101] The term "gene" refers to a coding region operably joined to appropriate regulatory sequences capable of regulating the expression of the polypeptide in some manner. A gene includes untranslated regulatory regions of DNA (*e.g.*, promoters, enhancers, repressors, etc.) preceding (upstream) and following (downstream) the coding region (open reading frame, ORF) as well as, where applicable, intervening sequences (*i.e.*, introns) between individual coding regions (*i.e.*, exons).

[000102] "Polypeptide" as used herein refers to any peptide, oligopeptide, polypeptide, gene product, expression product, or protein. A polypeptide is comprised of consecutive amino acids. The term "polypeptide" encompasses naturally occurring or synthetic molecules.

[000103] In addition, as used herein, the term "polypeptide" refers to amino acids joined to each other by peptide bonds or modified peptide bonds, *e.g.*, peptide isosteres, and may contain modified amino acids other than the 20 gene-encoded amino acids. The polypeptides can be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification can be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide can have many types of modifications. Modifications include, without



limitation, acetylation, acylation, ADP-ribosylation, amidation, covalent cross-linking or cyclization, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, disulfide bond formation,

5 demethylation, formation of cysteine or pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pergylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. (See *Proteins – Structure and Molecular Properties* 2nd Ed., T.E. Creighton, 10 W.H. Freeman and Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)).

[000104] As used herein, the term "amino acid sequence" refers to a list of abbreviations, letters, characters or words representing amino acid residues.

[000105] As used herein, the term "isolated" means that a material has been removed from 15 its original environment. For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides can be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and would be isolated in that such a vector or composition is not part 20 of its original environment.

[000106] As used herein with respect to nucleic acids, the term "recombinant" means that the nucleic acid is covalently joined and adjacent to a nucleic acid to which it is not adjacent in its natural environment. Additionally, as used herein with respect to a particular nucleic acid in a population of nucleic acids, the term "enriched" means that the nucleic acid 25 represents 5% or more of the number of nucleic acids in the population of molecules. Typically, the enriched nucleic acids represent 15% or more of the number of nucleic acids in the population of molecules. More typically, the enriched nucleic acids represent 50%, 90% or more of the number of nucleic acids in the population molecules.

[000107] "Recombinant" polypeptides or proteins refer to polypeptides or proteins 30 produced by recombinant DNA techniques, *i.e.*, produced from cells transformed by an exogenous recombinant DNA construct encoding the desired polypeptide or protein.

"Synthetic" polypeptides or proteins are those prepared by chemical synthesis (*e.g.*, solid-phase peptide synthesis). Chemical peptide synthesis is well known in the art (see, *e.g.*,



Merrifield (1963), *Am. Chem. Soc.* 85:2149-2154; Geysen et al. (1984), *Proc. Natl. Acad. Sci., USA* 81:3998) and synthesis kits and automated peptide synthesizer are commercially available (e.g., Cambridge Research Biochemicals, Cleveland, United Kingdom; Model 431A synthesizer from Applied Biosystems, Inc., Foster City, CA). Such equipment provides ready  
 5 access to the peptides of the invention, either by direct synthesis or by synthesis of a series of fragments that can be coupled using other known techniques.

[000108] As used herein with respect to pairs of nucleic acid or amino acid sequences, "identity" refers to the extent to which the two sequences are invariant at positions within the sequence which can be aligned. The percent identity between two given sequences can be  
 10 calculated using an algorithm such as BLAST (Altschul et al. (1990), *J. Mol. Biol.* 215:403-410).

When using the BLAST algorithm for sequences no longer than 250 nucleotides or about 80 amino acids ("short queries"), the search parameters can be as follows: the filter is off, the scoring matrix is PAM30, the word size is 3 or 2, the E value is 1000 or more, and the gap costs are 11, 1. For sequences longer  
 15 than 250 nucleotides or 80 amino acid residues, the default search parameters can be used. The BLAST website provides advice for special circumstances which is to be followed in such circumstances.

[000109] As used herein, "homology" has the same meaning as "identity" in the context of nucleotide sequences. However, with respect to amino acid sequences, "homology" includes  
 20 the percentage of identical and conservative amino acid substitutions. Percentages of homology can be calculated according to the algorithms of Smith and Waterman (1981), *Adv. Appl. Math.* 2:482.

[000110] As used herein in the context of two or more nucleic acid sequences, two sequences are "substantially identical" when they have at least 99.5% nucleotide identity,  
 25 when compared and aligned for maximum correspondence, as measured using the known sequence comparison algorithms described above. In addition, for purposes of determining whether sequences are substantially identical, synonymous codons in a coding region may be treated as identical to account for the degeneracy of the genetic code. Typically, the region for determination of substantial identity must span at least about 20 residues, and most  
 30 commonly the sequences are substantially identical over at least about 25-200 residues.

[000111] As used herein in the context of two or more amino acid sequences, two sequences are "substantially identical" when they have at least 99.5% identity, when compared and aligned for maximum correspondence, as measured using the known sequence



comparison algorithms described above. In addition, for purposes of determining whether sequences are substantially identical, conservative amino acid substitutions may be treated as identical if the polypeptide substantially retains its biological function.

[000112] "Hybridization" refers to the process by which a nucleic acid strand joins with a complementary strand through hydrogen bonding at complementary bases. Hybridization assays can be sensitive and selective so that a particular sequence of interest can be identified even in samples in which it is present at low concentrations. Stringent conditions are defined by concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. Stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature. In particular, as used herein, "stringent hybridization conditions" include 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 ng/ml sheared and denatured salmon sperm DNA, and equivalents thereof. Variations on the above ranges and conditions are well known in the art.

[000113] The term "variant" refers to polynucleotides or polypeptides of the invention modified at one or more nucleotides or amino acid residues (respectively) and wherein the encoded polypeptide or polypeptide retains nitrilase activity. Variants can be produced by any number of means including, for example, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site-saturated mutagenesis or any combination thereof.

[000114] Methods of making peptidomimetics based upon a known sequence is described, for example, in U.S. Patent Nos. 5,631,280; 5,612,895; and 5,579,250. Use of peptidomimetics can involve the incorporation of a non-amino acid residue with non-amide linkages at a given position. One aspect of the present invention is a peptidomimetic wherein the compound has a bond, a peptide backbone or an amino acid component replaced with a suitable mimic. Examples of unnatural amino acids which may be suitable amino acid mimics include  $\beta$ -alanine, L- $\alpha$ -amino butyric acid, L- $\gamma$ -amino butyric acid, L- $\alpha$ -amino isobutyric acid, L- $\epsilon$ -amino caproic acid, 7-amino heptanoic acid, L-aspartic acid, L-glutamic acid, N- $\epsilon$ -Boc-N- $\alpha$ -CBZ-L-lysine, N- $\epsilon$ -Boc-N- $\alpha$ -Fmoc-L-lysine, L-methionine sulfone, L-norleucine, L-norvaline, N- $\alpha$ -Boc-N- $\delta$ -CBZ-L-ornithine, N- $\delta$ -Boc-N- $\alpha$ -CBZ-L-ornithine, Boc-p-nitro-L-phenylalanine, Boc-hydroxyproline, Boc-L-thioprolin.



[000115] As used herein, "small molecule" encompasses a molecule having a molecular weight from about 20 daltons to about 1.5 kilodaltons.

[000116] The molecular biological techniques, such as subcloning, were performed using routine methods which would be well known to one of skill in the art. (Sambrook, J. Fritsch, EF, Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd ed.), Cold Spring Harbor Laboratory Press, Plainview NY.).

#### Computer Systems

[000117] In one aspect of the invention, any nucleic acid sequence and/or polypeptide sequence of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. Another aspect of the invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15 or 20 nucleic acid sequences as set forth in SEQ ID NOS: 1-386, and sequences substantially identical thereto. In a further aspect, another aspect is the comparison among and between nucleic acid sequences or polypeptide sequences of the invention and the comparison between sequences of the invention and other sequences by a computer. Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

[000118] Aspects of the invention include systems (e.g., internet based systems), particularly computer systems which store and manipulate the sequence information described herein. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze a sequence (either nucleic acid or polypeptide) as set forth in at least any one of SEQ ID NOS: 1-386 and sequences substantially identical thereto. The computer system typically includes a processor for processing, accessing and manipulating the sequence data. The processor can be any well-known type of central processing unit, such as, for example, the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq, AMD or International Business Machines.



[000119] Typically the computer system is a general purpose system that comprises the processor and one or more internal data storage components for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components.

[000120] In one particular aspect, the computer system includes a processor connected to a bus which is connected to a main memory (preferably implemented as RAM) and one or more internal data storage devices, such as a hard drive and/or other computer readable media having data recorded thereon. In some aspects, the computer system further includes one or more data retrieving device for reading the data stored on the internal data storage devices.

[000121] The data retrieving device may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, or a modem capable of connection to a remote data storage system (*e.g.*, via the internet) etc. In some aspects, the internal data storage device is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

[000122] The computer system includes a display which is used to display output to a computer user. It should also be noted that the computer system can be linked to other computer systems in a network or wide area network to provide centralized access to the computer system. In some aspects, the computer system may further comprise a sequence comparison algorithm. A "sequence comparison algorithm" refers to one or more programs which are implemented (locally or remotely) on the computer system to compare a nucleotide sequence with other nucleotide sequences and/or compounds stored within a data storage means.

#### Uses of Nitrilases

[000123] Nitrilases have been identified as key enzymes for the production of chiral  $\alpha$ -hydroxy acids, which are valuable intermediates in the fine chemicals industry, and as pharmaceutical intermediates. The nitrilase enzymes of the invention are useful to catalyze the stereoselective hydrolysis of cyanohydrins and aminonitriles, producing chiral  $\alpha$ -hydroxy- and  $\alpha$ -amino acids, respectively.



[000124] Stereoselective enzymes provide a key advantage over chemical resolution methods as they do not require harsh conditions and are more environmentally compatible. The use of nitrilases is of particular interest for the production of chiral amino acids and  $\alpha$ -hydroxy acids. Using a stereoselective nitrilase, dynamic resolution conditions can be established, due to the racemisation of the substrate under aqueous conditions. Thus 100% theoretical yields are achievable.

[000125] This invention is directed to the nitrilases which have been discovered and isolated from naturally occurring sources. This invention is also directed to evolving novel genes and gene pathways from diverse and extreme environmental sources. In an effort to develop the most extensive assortment of enzymes available, DNA was extracted directly from samples that have been collected from varying habitats around the globe. From these efforts, the largest collection of environmental genetic libraries in the world was developed. Through extensive high-throughput screening of these libraries, 192 new sequence-unique nitrilase enzymes have been discovered to date. Previous to this invention, fewer than 20 microbial-derived nitrilases had been reported in the literature and public databases.

[000126] Biocatalysts, such as nitrilases, play an important role in catalyzing metabolic reactions in living organisms. In addition, biocatalysts have found applications in the chemical industry, where they can perform many different reactions. Some examples of the advantages of the use of nitrilases is that they provide: high enantio-, chemo- and regio-selectivity; they function under mild reaction conditions; they provide direct access to products – with minimal protection; they have high catalytic efficiencies; they produce reduced waste compared with the chemical alternatives; they are easily immobilized as enzymes or cells; they are recoverable, recyclable and are capable of being manipulated via molecular biological techniques; they can be regenerated in whole cell processes; they are tolerant to organic solvents; and importantly, they can be evolved or optimized. Optimized nitrilases are presented herein as working examples of the invention.

[000127] Nitrilases catalyze the hydrolysis of nitrile moieties generating the corresponding carboxylic acid. Conventional chemical hydrolysis of nitriles requires strong acid or base and high temperature. However, one advantage of the invention is that nitrilases are provided which perform this reaction under mild conditions. Wide ranges of nitrile substrates can be transformed by nitrilases with high enantio-, chemo- and regio- selectivity.





Table 1 - Some characteristics of Nitrilases of the Invention

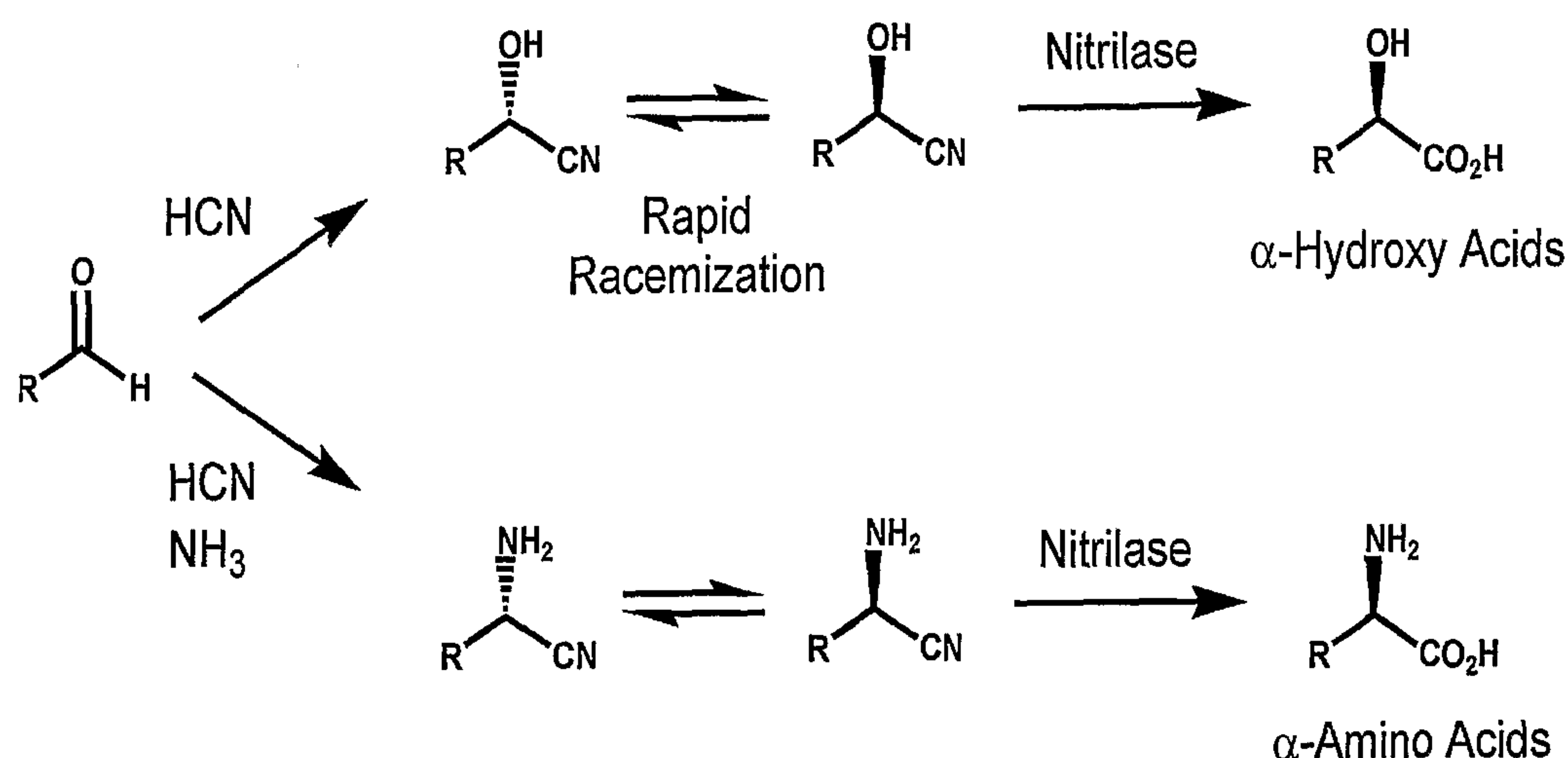
| <u>Previously Discovered Nitrilases</u> |                                                                             | <u>New Nitrilases</u>                                              |
|-----------------------------------------|-----------------------------------------------------------------------------|--------------------------------------------------------------------|
| Limitations                             | New Features                                                                | Benefits                                                           |
| < 20 reported                           | >180 newly discovered                                                       | Access to a wider substrate range                                  |
| Homologous                              | Unique nitrilases, many with little homology to previously known nitrilases |                                                                    |
| Narrow substrate Activity spectrum      | Broad substrate activity spectrum                                           |                                                                    |
| Very few shown to be enantioselective   | Enantioselective; both enantiomers accessible                               | Product with high enantiomeric excess and minimal waste production |
| Limited stability profile               | Stable in a variety of conditions                                           | Potential use in a wide range of process conditions                |
| Inconsistent supply                     | Consistent supply                                                           | Reliable source of product                                         |
| Not applicable                          | Amenable to optimization                                                    | Good source material leads to better product                       |

5

[000128] Dynamic Kinetic Resolution: The use of the nitrilases allows discrimination between two rapidly equilibrating enantiomers to give a single product in 100% theoretical yield. Nitrilases are utilized for dynamic resolution of key cyanohydrins and aminonitriles to produce enantiomerically pure  $\alpha$ -carboxylic and  $\alpha$ -amino acids. Newly discovered nitrilases disclosed herein yield products with >95% enantiomeric excess (*ee*) with and >95% yield. The nitrilases perform this transformation efficiently under mild conditions in aqueous solution or in the presence of organic solvent.

10





[000129] These products shown above also include the opposite enantiomers, although they are not shown. In one aspect, the invention provides an isolated nucleic acid having a sequence as set forth in any one of the Group A nucleic acid sequences, having a sequence substantially identical thereto, or having a sequence complementary thereto.

[000130] In another aspect, the invention provides an isolated nucleic acid including at least 20 consecutive nucleotides identical to a portion of a nucleotide sequence as set forth in the Group A nucleic acid sequences, having a sequence substantially identical thereto, or having a sequence complementary thereto.

[000131] In another aspect, the invention provides an isolated nucleic acid encoding a polypeptide having a sequence as set forth in the Group B amino acid sequences, or having a sequence substantially identical thereto.

[000132] In another aspect, the invention provides an isolated nucleic acid encoding a polypeptide having at least 10 consecutive amino acids identical to a portion of a sequence as set forth in the Group B amino acid sequences, or having a sequence substantially identical thereto.

[000133] In yet another aspect, the invention provides a substantially purified polypeptide comprising consecutive amino acid residues having a sequence as set forth in the Group B amino acid sequences, or having a sequence substantially identical thereto.



[000134] In another aspect, the invention provides an isolated antibody that specifically binds to a polypeptide of the invention. The invention also provides for a fragment of the antibody which retains the ability to specifically bind the polypeptide.

5 [000135] In another aspect, the invention provides a method of producing a polypeptide having a sequence as set forth in the Group B amino acid sequences, and sequences substantially identical thereto. The method includes introducing a nucleic acid encoding the polypeptide into a host cell, wherein the nucleic acid is operably joined to a promoter, and culturing the host cell under conditions that allow expression of the nucleic acid.

10 [000136] In another aspect, the invention provides a method of producing a polypeptide having at least 10 consecutive amino acids from a sequence as set forth in the Group B amino acid sequences, and sequences substantially identical thereto. The method includes introducing a nucleic acid encoding the polypeptide into a host cell, wherein the nucleic acid is operably joined to a promoter, and culturing the host cell under conditions that allow expression of the nucleic acid, thereby producing the polypeptide.

15 [000137] In another aspect, the invention provides a method of generating a variant of a nitrilase, including choosing a nucleic acid sequence as set forth in the Group A nucleic acid sequences, and changing one or more nucleotides in the sequence to another nucleotide, deleting one or more nucleotides in the sequence, or adding one or more nucleotides to the sequence.

20 [000138] In another aspect, the invention provides assays for identifying functional variants of the Group B amino acid sequences that retain the enzymatic function of the polypeptides of the Group B amino acid sequences. The assays include contacting a polypeptide comprising consecutive amino acid residues having a sequence identical to a sequence of the Group B amino acid sequences or a portion thereof, having a sequence substantially identical  
25 to a sequence of the Group B amino acid sequences or a portion thereof, or having a sequence which is a variant of a sequence of the Group B amino acid sequences that retains nitrilase activity, with a substrate molecule under conditions which allow the polypeptide to function, and detecting either a decrease in the level of substrate or an increase in the level of a specific reaction product of the reaction between the polypeptide and the substrate, thereby  
30 identifying a functional variant of such sequences.



Modification of Polypeptides of the Invention

[000139] Enzymes are highly selective catalysts. Their hallmark is the ability to catalyze reactions with exquisite stereo-selectivity, regio-selectivity, and chemo-selectivity that is unparalleled in conventional synthetic chemistry. Moreover, enzymes are remarkably versatile. They can be tailored to function in organic solvents, operate at extreme pHs (for example, acidic or basic conditions) extreme temperatures (for example, high temperatures and low temperatures), extreme salinity levels (for example, high salinity and low salinity), and catalyze reactions with compounds that can be structurally unrelated to their natural, physiological substrates except for the enzymatic active site.

[000140] The invention provides methods for modifying polypeptides having nitrilase activity or polynucleotides encoding such polypeptides in order to obtain new polypeptides which retain nitrilase activity but which are improved with respect to some desired characteristic. Such improvements can include the ability to function (*i.e.*, exhibit nitrilase activity) in organic solvents, operate at extreme or uncharacteristic pHs, operate at extreme or uncharacteristic temperatures, operate at extreme or uncharacteristic salinity levels, catalyze reactions with different substrates, etc.

[000141] The present invention directed to methods of using nitrilases so as to exploit the unique catalytic properties of these enzymes. Whereas the use of biocatalysts (*i.e.*, purified or crude enzymes) in chemical transformations normally requires the identification of a particular biocatalyst that reacts with a specific starting compound, the present invention uses selected biocatalysts and reaction conditions that are specific for functional groups that are present in many starting compounds. Each biocatalyst is specific for one functional group, or several related functional groups, and can react with many starting compounds containing this functional group.

[000142] Enzymes react at specific sites within a starting compound without affecting the rest of the molecule, a process which is very difficult to achieve using traditional chemical methods. This high degree of specificity provides the means to identify a single active compound within a library of compounds. The library is characterized by the series of biocatalytic reactions used to produce it, a so-called "biosynthetic history." Screening the library for biological activities and tracing the biosynthetic history identifies the specific



reaction sequence producing the active compound. The reaction sequence is repeated and the structure of the synthesized compound determined. This mode of identification, unlike other synthesis and screening approaches, does not require immobilization technologies, and compounds can be synthesized and tested free in solution using virtually any type of screening assay. It is important to note, that the high degree of specificity of enzyme reactions on functional groups allows for the "tracking" of specific enzymatic reactions that make up the biocatalytically produced library. (For further teachings on modification of molecules, including small molecules, see PCT Application Publication No. WO 1995/005475).

10 [000143] In one exemplification, the invention provides for the chimerization of a family of related nitrilase genes and their encoded family of related products. Thus according to this aspect of the invention, the sequences of a plurality of nitrilase nucleic acids (*e.g.*, the Group A nucleic acids) serve as nitrilase "templates" which are aligned using a sequence comparison algorithm such as those described above. One or more demarcation points are then identified  
15 in the aligned template sequences, which are located at one or more areas of homology. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks, which are used to generate chimeric nitrilases. Thus, the demarcation points identified and selected in the nitrilase template molecules serve as potential chimerization points in the assembly of the chimeric nitrilase molecules.

20 [000144] Typically, a useful demarcation point is an area of local identity between at least two progenitor templates, but preferably the demarcation point is an area of identity that is shared by at least half of the templates, at least two thirds of the templates, at least three fourths of the templates, or at nearly all of the templates.

[000145] The building blocks, which are defined by the demarcation points, can then be  
25 mixed (either literally, in solution, or theoretically, on paper or in a computer) and reassembled to form chimeric nitrilase genes. In one aspect, the gene reassembly process is performed exhaustively in order to generate an exhaustive library of all possible combinations. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same  
30 time, however, the order of assembly of each building block in the 5' to 3' direction in each combination is designed to reflect the order in the templates, and to reduce the production of unwanted, inoperative products.



[000146] In some aspects, the gene reassembly process is performed systematically, in order to generate a compartmentalized library with compartments that can be screened systematically, *e.g.*, one by one. In other words, the invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of chimeric products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, it allows a potentially very large number of chimeric molecules to be examined systematically in smaller groups.

[000147] In some aspects, the synthetic nature of the step in which the building blocks are generated or reassembled allows the design and introduction of sequences of nucleotides (*e.g.*, codons or introns or regulatory sequences) that can later be optionally removed in an *in vitro* process (*e.g.*, by mutagenesis) or in an *in vivo* process (*e.g.*, by utilizing the gene splicing ability of a host organism). The introduction of these nucleotides may be desirable for many reasons, including the potential benefit of creating a useful demarcation point.

[000148] The synthetic gene reassembly method of the invention utilizes a plurality of nucleic acid building blocks, each of which has two ligatable ends. Some examples of the two ligatable ends on each nucleic acid building block includes, but are not limited to, two blunt ends, or one blunt end and one overhang, or two overhangs. In a further, non-limiting example, the overhang can include one base pair, 2 base pairs, 3 base pairs, 4 base pairs or more.

[000149] A double-stranded nucleic acid building block can be of variable size. Preferred sizes for building blocks range from about 1 base pair (bp) (not including any overhangs) to about 100,000 base pairs (not including any overhangs). Other preferred size ranges are also provided, which have lower limits of from about 1 bp to about 10,000 bp (including every integer value in between), and upper limits of from about 2 bp to about 100,000 bp (including every integer value in between).

[000150] According to one aspect, a double-stranded nucleic acid building block is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that



form an overhang; thus containing no mismatches, apart from any overhang(s).

Alternatively, the two strands of a double-stranded nucleic acid building block can be complementary at fewer than every nucleotide, apart from any overhang(s). In particular, mismatches between the strands can be used to introduce codon degeneracy using methods  
5 such as the site-saturation mutagenesis described herein.

[000151] *In vivo* shuffling of molecules is also useful in providing variants and can be performed utilizing the natural property of cells to recombine multimers. While recombination *in vivo* has provided the major natural route to molecular diversity, genetic recombination remains a relatively complex process that involves (1) the recognition of  
10 homologues; (2) strand cleavage, strand invasion, and metabolic steps leading to the production of recombinant chiasma; and finally (3) the resolution of chiasma into discrete recombined molecules. The formation of the chiasma requires the recognition of homologous sequences.

[000152] Thus, the invention includes a method for producing a chimeric or recombinant  
15 polynucleotide from at least a first polynucleotide and a second polynucleotide *in vivo*. The invention can be used to produce a recombinant polynucleotide by introducing at least a first polynucleotide and a second polynucleotide which share at least one region of partial sequence homology (*e.g.*, the Group A nucleic acid sequences, and combinations thereof) into a suitable host cell. The regions of partial sequence homology promote processes which  
20 result in sequence reorganization producing a recombinant polynucleotide. Such hybrid polynucleotides can result from intermolecular recombination events which promote sequence integration between DNA molecules. In addition, such hybrid polynucleotides can result from intramolecular reductive reassortment processes which utilize repeated sequences to alter a nucleotide sequence within a DNA molecule.

[000153] The invention provides a means for generating recombinant polynucleotides  
25 which encode biologically active variant polypeptides (*e.g.*, a nitrilase variant). For example, a polynucleotide may encode a particular enzyme from one microorganism. An enzyme encoded by a first polynucleotide from one organism can, for example, function effectively under a particular environmental condition, *e.g.*, high salinity. An enzyme encoded by a  
30 second polynucleotide from a different organism can function effectively under a different environmental condition, such as extremely high temperature. A recombined polynucleotide containing sequences from the first and second original polynucleotides encodes a variant



enzyme which exhibits characteristics of both enzymes encoded by the original polynucleotides. Thus, the enzyme encoded by the recombined polynucleotide can function effectively under environmental conditions shared by each of the enzymes encoded by the first and second polynucleotides, *e.g.*, high salinity and extreme temperatures.

5 [000154] A variant polypeptide can exhibit specialized enzyme activity not displayed in the original enzymes. For example, following recombination and/or reductive reassortment of polynucleotides encoding nitrilase activity, the resulting variant polypeptide encoded by a recombined polynucleotide can be screened for specialized nitrilase activity obtained from each of the original enzymes, *i.e.*, the temperature or pH at which the nitrilase functions.

10 Sources of the original polynucleotides may be isolated from individual organisms ("isolates"), collections of organisms that have been grown in defined media ("enrichment cultures"), or, uncultivated organisms ("environmental samples"). The use of a culture-independent approach to derive polynucleotides encoding novel bioactivities from environmental samples is most preferable since it allows one to access untapped resources of

15 biodiversity. The microorganisms from which the polynucleotide may be prepared include prokaryotic microorganisms, such as *Xanthobacter*, *Eubacteria* and *Archaeobacteria*, and lower eukaryotic microorganisms such as fungi, some algae and protozoa. Polynucleotides may be isolated from environmental samples in which case the nucleic acid may be recovered without culturing of an organism or recovered from one or more cultured organisms. In one

20 aspect, such microorganisms may be extremophiles, such as hyperthermophiles, psychrophiles, psychrotrophs, halophiles, barophiles and acidophiles. Polynucleotides encoding enzymes isolated from extremophilic microorganisms are particularly preferred. Such enzymes may function at temperatures above 100°C in terrestrial hot springs and deep sea thermal vents, at temperatures below 0°C in arctic waters, in the saturated salt

25 environment of the Dead Sea, at pH values around 0 in coal deposits and geothermal sulfur-rich springs, or at pH values greater than 11 in sewage sludge.

[000155] Examples of mammalian expression systems that can be employed to express recombinant proteins include the COS-7, C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors comprise an origin of replication, a suitable promoter and

30 enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice and polyadenylation sites may be



used to provide the required nontranscribed genetic elements.

[000156] Host cells containing the polynucleotides of interest can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan. Clones, which are identified as having a desired enzyme activity or other property may then be sequenced to identify the recombinant polynucleotide sequence encoding the enzyme having the desired activity or property.

10 [000157] In one aspect, the invention provides for the isolated nitrilases as either an isolated nucleic acid or an isolated polypeptide wherein the nucleic acid or the polypeptide was prepared by recovering DNA from a DNA population derived from at least one uncultivated microorganism, and transforming a host with recovered DNA to produce a library of clones which is screened for the specified protein, *e.g.* nitrilase activity. U.S. Patent No. 6,280,926,  
15 Short, provides descriptions of such methods.

[000158] Therefore, in a one aspect, the invention relates to a method for producing a biologically active recombinant nitrilase polypeptide and screening such a polypeptide for desired activity or property by:

- 20 1) introducing at least a first nitrilase polynucleotide and a second nitrilase polynucleotide, said at least first nitrilase polynucleotide and second nitrilase polynucleotide sharing at least one region of sequence homology, into a suitable host cell;
- 2) growing the host cell under conditions which promote sequence reorganization resulting in a recombinant nitrilase polynucleotide;
- 25 3) expressing a recombinant nitrilase polypeptide encoded by the recombinant nitrilase polynucleotide;
- 4) screening the recombinant nitrilase polypeptide for the desired activity or property; and
- 5) isolating the recombinant nitrilase polynucleotide encoding the recombinant  
30 nitrilase polypeptide.



[000159] Examples of vectors which may be used include viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral DNA (e.g., vaccinia, adenovirus, fowlpox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for the hosts of interest (e.g., *Bacillus*, *Aspergillus* and yeast). Large numbers of suitable vectors are known to those of skill in the art, and are commercially available.

Examples of bacterial vectors include pQE vectors (Qiagen, Valencia, CA); pBluescript plasmids, pNH vectors, and lambda-ZAP vectors (Stratagene, La Jolla, CA); and pTRC99a, pKK223-3, pDR540, and pRIT2T vectors (Pharmacia, Peapack, NJ). Examples of eukaryotic vectors include pXT1 and pSG5 vectors (Stratagene, La Jolla, CA); and pSVK3, pBPV, pMSG, and pSVLSV40 vectors (Pharmacia, Peapack, NJ). However, any other plasmid or other vector may be used so long as they are replicable and viable in the host.

[000160] A preferred type of vector for use in the present invention contains an f-factor (or fertility factor) origin of replication. The f-factor in *E. coli* is a plasmid which effects high frequency transfer of itself during conjugation and less frequent transfer of the bacterial chromosome itself. A particularly preferred aspect is to use cloning vectors referred to as "fosmids" or bacterial artificial chromosome (BAC) vectors. These are derived from *E. coli* f-factor which is able to stably integrate large segments of genomic DNA.

[000161] The DNA sequence in the expression vector is operably joined to appropriate expression control sequences, including a promoter, to direct RNA synthesis. Useful bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P<sub>R</sub>, P<sub>L</sub> and trp. Useful eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

[000162] In addition, the expression vectors can contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Useful selectable markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or tetracycline or ampicillin resistance in *E. coli*.



[000163] The vector may be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation

5 [000164] Reductive Reassortment - In another aspect, variant nitrilase polynucleotides can be generated by the process of reductive reassortment. Whereas recombination is an “inter-molecular” process which, in bacteria, is generally viewed as a “recA-dependent” phenomenon, the process of “reductive reassortment” occurs by an “intra-molecular”, recA-independent process. In this aspect, the invention can rely on the ability of cells to mediate  
10 reductive processes to decrease the complexity of quasi-repeated sequences in the cell by deletion. The method involves the generation of constructs containing consecutive repeated or quasi-repeated sequences (original encoding sequences), the insertion of these sequences into an appropriate vector, and the subsequent introduction of the vector into an appropriate host cell. The reassortment of the individual molecular identities occurs by combinatorial  
15 processes between the consecutive sequences in the construct possessing regions of homology, or between quasi-repeated units. The reassortment process recombines and/or reduces the complexity and extent of the repeated sequences, and results in the production of novel molecular species. Various treatments may be applied to enhance the rate of reassortment, such as ultra-violet light or DNA damaging chemicals. In addition, host cell  
20 lines displaying enhanced levels of “genetic instability” can be used.

[000165] Repeated Sequences - Repeated or “quasi-repeated” sequences play a role in genetic instability. In the present invention, “quasi-repeats” are repeats that are not identical in structure but, rather, represent an array of consecutive sequences which have a high degree of similarity or identity sequences. The reductive reassortment or deletion process in the cell  
25 reduces the complexity of the resulting construct by deleting sequences between positions within quasi-repeated sequences. Because the deletion (and potentially insertion) events can occur virtually anywhere within the quasi-repetitive units, these sequences provide a large repertoire of potential variants.

[000166] When the quasi-repeated sequences are all ligated in the same orientation, for  
30 instance head-to-tail or vice versa, the endpoints of a deletion are, for the most part, equally likely to occur anywhere within the quasi-repeated sequences. In contrast, when the units are presented head-to-head or tail-to-tail, the inverted quasi-repeated sequences can form a



duplex which delineates the endpoints of the adjacent units and thereby favors deletion of discrete units. Therefore, it is preferable in the present invention that the quasi-repeated sequences are joined in the same orientation because random orientation of quasi-repeated sequences will result in the loss of reassortment efficiency, while consistent orientation of the sequences will offer the highest efficiency. Nonetheless, although having fewer of the contiguous sequences in the same orientation decreases the efficiency or reductive reassortment, it may still provide sufficient variation for the effective recovery of novel molecules.

[000167] Sequences can be assembled in a head-to-tail orientation using any of a variety of methods, including the following:

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

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

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

[000168] The recovery of the reassorted sequences relies on the identification of cloning vectors with a reduced repetitive index (RI). The reassorted coding sequences can then be recovered by amplification. The products are recloned and expressed. The recovery of cloning vectors with reduced RI can be effected by:

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

2) The physical recovery of shortened vectors by physical procedures. In this case, the cloning vector would be recovered using standard plasmid isolation procedures and then size-fractionated using standard procedures (*e.g.*, agarose gel or column with a low molecular weight cut off).

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

4) The use of direct selection techniques wherein an expression vector is used and the appropriate selection is carried out.



[000169] Coding sequences from related organisms may demonstrate a high degree of homology but nonetheless encode quite diverse protein products. These types of sequences are particularly useful in the present invention as quasi-repeats. However, while the examples illustrated below demonstrate the reassortment of coding sequences with a high degree of identity (quasi-repeats), this process is not limited to nearly identical repeats.

[000170] The following example demonstrates a method of the invention. Quasi-repetitive coding sequences derived from three different species are obtained. Each sequence encodes a protein with a distinct set of properties. Each of the sequences differs by one or more base pairs at unique positions in the sequences which are designated "A", "B" and "C". The quasi-repeated sequences are separately or collectively amplified and ligated into random assemblies such that all possible permutations and combinations are available in the population of ligated molecules. The number of quasi-repeat units can be controlled by the assembly conditions. The average number of quasi-repeated units in a construct is defined as the repetitive index (RI).

[000171] Once formed, the constructs can be size-fractionated on an agarose gel according to published protocols, inserted into a cloning vector, and transfected into an appropriate host cell. The cells can then be propagated to allow reductive reassortment to occur. The rate of the reductive reassortment process may be stimulated by the introduction of DNA damage if desired. Whether the reduction in RI is mediated by deletion formation between repeated sequences by an "intra-molecular" mechanism, or mediated by recombination-like events through "inter-molecular" mechanisms is immaterial. The end result is a reassortment of the molecules into all possible combinations.

[000172] In another aspect, prior to or during recombination or reassortment, polynucleotides of the invention or polynucleotides generated by the methods described herein can be subjected to agents or processes which promote the introduction of mutations into the original polynucleotides. The introduction of such mutations would increase the diversity of resulting hybrid polynucleotides and polypeptides encoded therefrom. The agents or processes which promote mutagenesis include, but are not limited to: (+)-CC-1065, or a synthetic analog such as (+)-CC-1065-(N3-adenine) (Sun et al. (1992), *Biochemistry* 31(10):2822-9); an N-acetylated or deacetylated 4'-fluoro-4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see, for example, van de Poll et al. (1992), *Carcinogenesis* 13(5):751-8); or a N-acetylated or deacetylated 4-aminobiphenyl adduct capable of inhibiting



DNA synthesis (see also, Van de Poll et al. (1992), *supra*); trivalent chromium, a trivalent chromium salt, a polycyclic aromatic hydrocarbon ("PAH") DNA adduct capable of inhibiting DNA replication, such as 7-bromomethyl-benz[a]anthracene ("BMA"), tris(2,3-dibromopropyl)phosphate ("Tris-BP"), 1,2-dibromo-3-chloropropane ("DBCP"), 2-bromoacrolein (2BA), benzo[a]pyrene-7,8-dihydrodiol-9-10-epoxide ("BPDE"), a platinum(II) halogen salt, N-hydroxy-2-amino-3-methylimidazo[4,5-f]-quinoline ("N-hydroxy-IQ"), and N-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-f]-pyridine ("N-hydroxy-PhIP"). Especially preferred means for slowing or halting PCR amplification consist of UV light (+)-CC-1065 and (+)-CC-1065-(N3-Adenine). Particularly encompassed means are DNA adducts or polynucleotides comprising the DNA adducts from the polynucleotides or polynucleotides pool, which can be released or removed by a process including heating the solution comprising the polynucleotides prior to further processing.

[000173] GSSM<sup>TM</sup> - The invention also provides for the use of codon primers containing a degenerate N,N,G/T sequence to introduce point mutations into a polynucleotide, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is represented at each amino acid position, a method referred to as gene site-saturated mutagenesis (GSSM<sup>TM</sup>). The oligos used are comprised contiguously of a first homologous sequence, a degenerate N,N,G/T sequence, and possibly a second homologous sequence. The progeny translational products from the use of such oligos include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,G/T sequence includes codons for all 20 amino acids.

[000174] In one aspect, one such degenerate oligo (comprising one degenerate N,N,G/T cassette) is used for subjecting each original codon in a parental polynucleotide template to a full range of codon substitutions. In another aspect, at least two degenerate N,N,G/T cassettes are used – either in the same oligo or not, for subjecting at least two original codons in a parental polynucleotide template to a full range of codon substitutions. Thus, more than one N,N,G/T sequence can be contained in one oligo to introduce amino acid mutations at more than one site. This plurality of N,N,G/T sequences can be directly contiguous, or separated by one or more additional nucleotide sequences. In another aspect, oligos serviceable for introducing additions and deletions can be used either alone or in combination with the codons containing an N,N,G/T sequence, to introduce any combination or permutation of amino acid additions, deletions, and/or substitutions.



[000175] In a particular exemplification, it is possible to simultaneously mutagenize two or more contiguous amino acid positions using an oligo that contains contiguous N,N,G/T triplets, *i.e.*, a degenerate (N,N,G/T)<sub>n</sub> sequence.

5 [000176] In another aspect, the present invention provides for the use of degenerate cassettes having less degeneracy than the N,N,G/T sequence. For example, it may be desirable in some instances to use a degenerate triplet sequence comprised of only one N, where said N can be in the first second or third position of the triplet. Any other bases including any combinations and permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some instances to use a  
10 degenerate N,N,N triplet sequence, or an N,N, G/C triplet sequence.

[000177] It is appreciated, however, that the use of a degenerate triplet (such as N,N,G/T or an N,N, G/C triplet sequence) as disclosed in the instant invention is advantageous for several reasons. In one aspect, this invention provides a means to systematically and fairly easily generate the substitution of the full range of the 20 possible amino acids into each and every  
15 amino acid position in a polypeptide. Thus, for a 100 amino acid polypeptide, the invention provides a way to systematically and fairly easily generate 2000 distinct species (*i.e.*, 20 possible amino acids per position times 100 amino acid positions). It is appreciated that there is provided, through the use of an oligo containing a degenerate N,N,G/T or an N,N, G/C triplet sequence, 32 individual sequences that code for the 20 possible amino acids. Thus, in  
20 a reaction vessel in which a parental polynucleotide sequence is subjected to saturation mutagenesis using one such oligo, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligo in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel.

[000178] This invention also provides for the use of nondegenerate oligonucleotides, which  
25 can optionally be used in combination with degenerate primers disclosed. It is appreciated that in some situations, it is advantageous to use nondegenerate oligos to generate specific point mutations in a working polynucleotide. This provides a means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes, and point mutations that cause the generation of stop codons and the corresponding expression of  
30 polypeptide fragments.



[000179] Thus, in one aspect, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide molecules such that all 20 amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide. The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (*e.g.*, cloned into a suitable *E. coli* host using an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in property (when compared to the parental polypeptide), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

[000180] It is appreciated that upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid, and each of two favorable changes) and 3 positions. Thus, there are  $3 \times 3 \times 3$  or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (*i.e.*, 2 at each of three positions) and no change at any position.

[000181] In yet another aspect, site-saturation mutagenesis can be used together with shuffling, chimerization, recombination and other mutagenizing processes, along with screening. This invention provides for the use of any mutagenizing process(es), including saturation mutagenesis, in an iterative manner. In one exemplification, the iterative use of any mutagenizing process(es) is used in combination with screening.

[000182] Thus, in a non-limiting exemplification, polynucleotides and polypeptides of the invention can be derived by saturation mutagenesis in combination with additional mutagenization processes, such as process where two or more related polynucleotides are introduced into a suitable host cell such that a hybrid polynucleotide is generated by recombination and reductive reassortment.



[000183] In addition to performing mutagenesis along the entire sequence of a gene, mutagenesis can be used to replace each of any number of bases in a polynucleotide sequence, wherein the number of bases to be mutagenized can be each integer from about 15 to about 100,000. Thus, instead of mutagenizing every position along a molecule, one can  
5 subject every or a discrete number of bases (*e.g.*, a subset totaling from about 15 to about 100,000) to mutagenesis. In one aspect, a separate nucleotide is used for mutagenizing each position or group of positions along a polynucleotide sequence. A group of 3 positions to be mutagenized can be a codon. In one aspect, the mutations are introduced using a mutagenic primer, containing a heterologous cassette, also referred to as a mutagenic cassette. For  
10 example, cassettes can have from about 1 to about 500 bases. Each nucleotide position in such heterologous cassettes can be N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, or E, where E is any base that is not A, C, G, or T.

[000184] In a general sense, saturation mutagenesis comprises mutagenizing a complete set of mutagenic cassettes (for example, each cassette is about 1-500 bases in length) in a defined  
15 polynucleotide sequence to be mutagenized (for example, the sequence to be mutagenized is from about 15 to about 100,000 bases in length). Thus, a group of mutations (ranging from about 1 to about 100 mutations) is introduced into each cassette to be mutagenized. A grouping of mutations to be introduced into one cassette can be different or the same from a second grouping of mutations to be introduced into a second cassette during the application  
20 of one round of saturation mutagenesis. Such groupings are exemplified by deletions, additions, groupings of particular codons, and groupings of particular nucleotide cassettes.

[000185] Defined sequences to be mutagenized include a whole gene, pathway, cDNA, entire open reading frame (ORF), promoter, enhancer, repressor/transactivator, origin of replication, intron, operator, or any polynucleotide functional group. Generally, a “defined  
25 sequence” for this purpose may be any polynucleotide that a 15 base-polynucleotide sequence, and polynucleotide sequences of lengths between about 15 bases and about 15,000 bases (this invention specifically names every integer in between). Considerations in choosing groupings of codons include types of amino acids encoded by a degenerate mutagenic cassette.

[000186] In a particularly preferred exemplification a grouping of mutations that can be introduced into a mutagenic cassette, this invention specifically provides for degenerate codon substitutions (using degenerate oligos) that code for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,



14, 15, 16, 17, 18, 19, and 20 amino acids at each position, and a library of polypeptides encoded thereby.

[000187] One aspect of the invention is an isolated nucleic acid comprising one of the sequences of the Group A nucleic acid sequences, sequences substantially identical thereto, sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of the Group A nucleic acid sequences. The isolated nucleic acids may comprise DNA, including cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. Alternatively, the isolated nucleic acids may comprise RNA.

[000188] As discussed in more detail below, the isolated nucleic acid sequences of the invention may be used to prepare one of the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of one of the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto.

[000189] Alternatively, the nucleic acid sequences of the invention may be mutagenized using conventional techniques, such as site directed mutagenesis, or other techniques familiar to those skilled in the art, to introduce silent changes into the polynucleotides of the Group A nucleic acid sequences, and sequences substantially identical thereto. As used herein, "silent changes" include, for example, changes which do not alter the amino acid sequence encoded by the polynucleotide. Such changes may be desirable in order to increase the level of the polypeptide produced by host cells containing a vector encoding the polypeptide by introducing codons or codon pairs which occur frequently in the host organism.

[000190] The invention also relates to polynucleotides which have nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptides of the invention (*e.g.*, the Group B amino acid sequences). Such nucleotide changes may be introduced using techniques such as site-directed mutagenesis, random chemical mutagenesis, exonuclease III deletion, and other recombinant DNA techniques. Alternatively, such nucleotide changes may be naturally occurring allelic variants which are isolated by identifying nucleic acid sequences which specifically hybridize to probes



comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of the Group A nucleic acid sequences, and sequences substantially identical thereto (or the sequences complementary thereto) under conditions of high, moderate, or low stringency as provided herein.

5 Immobilized Enzyme Solid Supports

[000191] The enzymes, fragments thereof and nucleic acids which encode the enzymes and fragments can be affixed to a solid support. This is often economical and efficient in the use of the enzymes in industrial processes. For example, a consortium or cocktail of enzymes (or active fragments thereof), which are used in a specific chemical reaction, can be attached to a solid support and dunked into a process vat. The enzymatic reaction can occur. Then, the solid support can be taken out of the vat, along with the enzymes affixed thereto, for repeated use. In one aspect of the invention, the isolated nucleic acid is affixed to a solid support. In another aspect of the invention, the solid support is selected from the group of a gel, a resin, a polymer, a ceramic, a glass, a microelectrode and any combination thereof.

15 [000192] For example, solid supports useful in this invention include gels. Some examples of gels include sepharose, gelatin, glutaraldehyde, chitosan-treated glutaraldehyde, albumin-glutaraldehyde, chitosan-Xanthan, toyopearl gel (polymer gel), alginate, alginate-polylysine, carrageenan, agarose, glyoxyl agarose, magnetic agarose, dextran-agarose, poly(Carbamoyl Sulfonate) hydrogel, BSA-PEG hydrogel, phosphorylated polyvinyl alcohol (PVA),  
20 monoaminoethyl-N-aminoethyl (MANA), amino, or any combination thereof.

[000193] Another solid support useful in the present invention are resins or polymers. Some examples of resins or polymers include cellulose, acrylamide, nylon, rayon, polyester, anion-exchange resin, AMBERLITE™ XAD-7, AMBERLITE™ XAD-8, AMBERLITE™ IRA-94, AMBERLITE™ IRC-50, polyvinyl, polyacrylic, polymethacrylate, or any  
25 combination thereof. Another type of solid support useful in the present invention is ceramic. Some examples include non-porous ceramic, porous ceramic, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>. Another type of solid support useful in the present invention is glass. Some examples include non-porous glass, porous glass, aminopropyl glass or any combination thereof. Another type of solid support which can be used is a microelectrode. An example is a polyethyleneimine-coated  
30 magnetite. Graphitic particles can be used as a solid support. Another example of a solid support is a cell, such as a red blood cell.



Methods of immobilization

[000194] There are many methods which would be known to one of skill in the art for immobilizing enzymes or fragments thereof, or nucleic acids, onto a solid support. Some examples of such methods include electrostatic droplet generation, electrochemical means, via adsorption, via covalent binding, via cross-linking, via a chemical reaction or process, via encapsulation, via entrapment, via calcium alginate, or via poly (2-hydroxyethyl methacrylate). Like methods are described in *Methods in Enzymology, Immobilized Enzymes and Cells*, Part C. 1987. Academic Press. Edited by S. P. Colowick and N. O. Kaplan.

Volume 136; and *Immobilization of Enzymes and Cells*. 1997. Humana Press. Edited by G. F. Bickerstaff. Series: Methods in Biotechnology, Edited by J. M. Walker.

[000195] Probes - The isolated nucleic acids of the Group A nucleic acid sequences, sequences substantially identical thereto, complementary sequences, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the foregoing sequences may also be used as probes to determine whether a biological sample, such as a soil sample, contains an organism having a nucleic acid sequence of the invention or an organism from which the nucleic acid was obtained. In such procedures, a biological sample potentially harboring the organism from which the nucleic acid was isolated is obtained and nucleic acids are obtained from the sample. The nucleic acids are contacted with the probe under conditions which permit the probe to specifically hybridize to any complementary sequences which are present therein.

[000196] Where necessary, conditions which permit the probe to specifically hybridize to complementary sequences may be determined by placing the probe in contact with complementary sequences from samples known to contain the complementary sequence as well as control sequences which do not contain the complementary sequence. Hybridization conditions, such as the salt concentration of the hybridization buffer, the formamide concentration of the hybridization buffer, or the hybridization temperature, can be varied to identify conditions which allow the probe to hybridize specifically to complementary nucleic acids. Stringent hybridization conditions are recited herein.

[000197] Hybridization may be detected by labeling the probe with a detectable agent such as a radioactive isotope, a fluorescent dye or an enzyme capable of catalyzing the formation of a detectable product. Many methods for using the labeled probes to detect the presence of complementary nucleic acids in a sample are familiar to those skilled in the art. These include Southern Blots, Northern Blots, colony hybridization procedures, and dot blots.



Protocols for each of these procedures are provided in Ausubel et al. (1997), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., and Sambrook et al. (1989), *Molecular Cloning: A Laboratory Manual 2d Ed.*, Cold Spring Harbor Laboratory Press,

- 5 [000198] In one example, a probe DNA is “labeled” with one partner of a specific binding pair (*i.e.*, a ligand) and the other partner of the pair is bound to a solid matrix to provide ease of separation of target from its source. For example, the ligand and specific binding partner can be selected from, in either orientation, the following: (1) an antigen or hapten and an antibody or specific binding fragment thereof; (2) biotin or iminobiotin and avidin or streptavidin; (3) a sugar and a lectin specific therefor; (4) an enzyme and an inhibitor therefor; (5) an apoenzyme and cofactor; (6) complementary homopolymeric oligonucleotides; and (7) a hormone and a receptor therefor. In one example, the solid phase is selected from: (1) a glass or polymeric surface; (2) a packed column of polymeric beads; and (3) magnetic or paramagnetic particles.
- 10 [000199] Alternatively, more than one probe (at least one of which is capable of specifically hybridizing to any complementary sequences which are present in the nucleic acid sample), may be used in an amplification reaction to determine whether the sample contains an organism containing a nucleic acid sequence of the invention (*e.g.*, an organism from which the nucleic acid was isolated). Typically, the probes comprise oligonucleotides. In one aspect, the amplification reaction may comprise a PCR reaction. PCR protocols are described in Ausubel et al. (1997), *supra*, and Sambrook et al. (1989), *supra*. Alternatively, the amplification may comprise a ligase chain reaction, 3SR, or strand displacement reaction. (See Barany (1991), *PCR Methods and Applications* 1:5-16; Fahy et al. (1991), *PCR Methods and Applications* 1:25-33; and Walker et al. (1992), *Nucleic Acid Research* 20:1691-1696.
- 15 [000200] Probes derived from sequences near the ends of a sequence as set forth in the Group A nucleic acid sequences, and sequences substantially identical thereto, may also be used in chromosome walking procedures to identify clones containing genomic sequences located adjacent to the nucleic acid sequences as set forth above. Such methods allow the isolation of genes which encode additional proteins from the host organism.
- 20 [000201] An isolated nucleic acid sequence as set forth in the Group A nucleic acid sequences, sequences substantially identical thereto, sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or



500 consecutive bases of one of the foregoing sequences may be used as probes to identify and isolate related nucleic acids. In some aspects, the related nucleic acids may be cDNAs or genomic DNAs from organisms other than the one from which the nucleic acid was isolated. For example, the other organisms may be related organisms. In such procedures, a nucleic acid sample is contacted with the probe under conditions which permit the probe to specifically hybridize to related sequences. Hybridization of the probe to nucleic acids from the related organism is then detected using any of the methods described above.

[000202] In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized.

For example, the length of the nucleic acids, the amount of complementarity between the nucleic acids, the nucleotide sequence composition (*e.g.*, G-C rich v. A-T rich content), and the nucleic acid type (*e.g.*, RNA v. DNA) can be considered in selecting hybridization conditions. Stringency may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature,  $T_m$ , is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly complementary probe. Stringent conditions are selected to be equal to or about 5°C lower than the  $T_m$  for a particular probe. The melting temperature of the probe may be calculated using the following formulas:

[000203] For probes between 14 and 70 nucleotides in length the melting temperature ( $T_m$ ) is calculated using the formula:  $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$  where N is the length of the probe.

[000204] If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation:  $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$  where N is the length of the probe.

[000205] Expression Libraries - Expression libraries can be created using the polynucleotides of the invention in combination with expression vectors and appropriate host cells. The library allows for the *in vivo* expression of the polypeptides which are encoded by the polynucleotides of the invention. After such expression libraries have been generated one can include the additional step of "biopanning" such libraries prior to screening by cell sorting. The "biopanning" procedure refers to a process for identifying clones having a specified biological activity by screening for sequence identity in a library of clones prepared by (i) selectively isolating target DNA derived from at least one microorganism by use of at least one probe DNA comprising at least a portion of a DNA sequence encoding a



polypeptide having a specified biological activity (*e.g.*, nitrilase activity); and (ii) optionally transforming a host with the isolated target DNA to produce a library of clones which are screened for the specified biological activity.

[000206] The probe DNA used for selectively isolating the target DNA of interest from the

5 DNA derived from at least one microorganism can be a full-length coding region sequence or a partial coding region sequence of DNA for an enzyme of known activity. The original DNA library can be probed using mixtures of probes comprising at least a portion of DNA sequences encoding enzymes having the specified enzyme activity. These probes or probe libraries are single-stranded and the microbial DNA which is probed has been converted into  
10 single-stranded form. The probes that are particularly suitable are those derived from DNA encoding enzymes having an activity similar or identical to the specified enzyme activity that is to be screened.

[000207] Having prepared a multiplicity of clones from DNA selectively isolated from an organism, such clones are screened for a specific enzyme activity and to identify the clones  
15 having the specified enzyme characteristics.

[000208] The screening for enzyme activity may be affected on individual expression clones or may be initially affected on a mixture of expression clones to ascertain whether or not the mixture has one or more specified enzyme activities. If the mixture has a specified enzyme activity, then the individual clones may be rescreened for such enzyme activity or for  
20 a more specific activity. Thus, for example, if a clone mixture has nitrilase activity, then the individual clones may be recovered and screened to determine which of such clones has nitrilase activity.

[000209] As described with respect to one of the above aspects, the invention provides a process for enzyme activity screening of clones containing selected DNA derived from a  
25 microorganism which process includes: screening a library for specified enzyme activity, said library including a plurality of clones, said clones having been prepared by recovering from genomic DNA of a microorganism selected DNA, which DNA is selected by hybridization to at least one DNA sequence which is all or a portion of a DNA sequence encoding an enzyme having the specified activity; and transforming a host with the selected  
30 DNA to produce clones which are screened for the specified enzyme activity.

[000210] In one aspect, a DNA library derived from a microorganism is subjected to a selection procedure to select therefrom DNA which hybridizes to one or more probe DNA



sequences which is all or a portion of a DNA sequence encoding an enzyme having the specified enzyme activity by:

(a) contacting the single-stranded DNA population from the DNA library with the DNA probe bound to a ligand under stringent hybridization conditions so as to produce a duplex between the probe and a member of the DNA library;

(b) contacting the duplex with a solid phase specific binding partner for the ligand so as to produce a solid phase complex;

(c) separating the solid phase complex from the non-duplexed members of the DNA library;

(d) denaturing the duplex to release the member of the DNA library;

(e) creating a complementary DNA strand of the member from step (d) so as to make the member a double-stranded DNA;

(f) introducing the double-stranded DNA into a suitable host so as to express a polypeptide which is encoded by the member DNA; and

(g) determining whether the polypeptide expressed exhibits the specified enzymatic activity.

[000211] In another aspect, the process includes a preselection to recover DNA including signal or secretion sequences. In this manner it is possible to select from the genomic DNA population by hybridization as hereinabove described only DNA which includes a signal or secretion sequence. The following paragraphs describe the protocol for this aspect of the invention, the nature and function of secretion signal sequences in general and a specific exemplary application of such sequences to an assay or selection process.

[000212] A particularly aspect of this aspect further comprises, after (a) but before (b) above, the steps of:

(i) contacting the single-stranded DNA population of (a) with a ligand-bound oligonucleotide probe that is complementary to a secretion signal sequence unique to a given class of proteins under hybridization conditions to form a double-stranded DNA duplex;

(ii) contacting the duplex of (i) with a solid phase specific binding partner for said ligand so as to produce a solid phase complex;

(iii) separating the solid phase complex from the single-stranded DNA population of (a);

(iv) denaturing the duplex so as to release single-stranded DNA members of the genomic population; and



(v) separating the single-stranded DNA members from the solid phase bound probe.

[000213] The DNA which has been selected and isolated to include a signal sequence is then subjected to the selection procedure hereinabove described to select and isolate therefrom DNA which binds to one or more probe DNA sequences derived from DNA encoding an enzyme(s) having the specified enzyme activity. This procedure is described and exemplified in U.S. Pat. No. 6,054,267.

[000214] *In vivo* biopanning may be performed utilizing a (fluorescence activated cell sorter) FACS-based machine. Complex gene libraries are constructed with vectors which contain elements which stabilize transcribed RNA. For example, the inclusion of sequences which result in secondary structures such as hairpins which are designed to flank the transcribed regions of the RNA would serve to enhance their stability, thus increasing their half life within the cell. The probe molecules used in the biopanning process consist of oligonucleotides labeled with reporter molecules that only fluoresce upon binding of the probe to a target molecule. These probes are introduced into the recombinant cells from the library using one of several transformation methods. The probe molecules bind to the transcribed target mRNA resulting in DNA/RNA heteroduplex molecules. Binding of the probe to a target will yield a fluorescent signal that is detected and sorted by the FACS machine during the screening process.

[000215] In some aspects, the nucleic acid encoding one of the polypeptides of the Group B amino acid sequences, sequences substantially identical thereto, or fragments comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptide or fragment thereof. Optionally, the nucleic acid can encode a fusion polypeptide in which one of the polypeptides of the Group B amino acid sequences, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is fused to heterologous peptides or polypeptides, such as N-terminal identification peptides which impart desired characteristics, such as increased stability or simplified purification.

[000216] The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells, eukaryotic cells, mammalian cells, insect cells, or plant cells. As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, fungal cells, such as yeast,



insect cells such as *Drosophila* S2 and *Spodoptera* Sf9, animal cells such as CHO, COS or Bowes melanoma, and adenoviruses. The selection of an appropriate host is within the abilities of those skilled in the art.

[000217] Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (*e.g.*, temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or fragment thereof.

[000218] Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

[000219] Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts (described by Gluzman (1981), *Cell* 23:175.), and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines.

[000220] The invention also relates to variants of the polypeptides of the Group B amino acid sequences, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. In particular, the variants may differ in amino acid sequence from the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto, by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.



[000221] The variants may be naturally occurring or created *in vitro*. In particular, such variants may be created using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives  
 5 may be created using chemical synthesis or modification procedures.

[000222] Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids which encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large  
 10 number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Typically, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

#### Error Prone PCR

15 [000223] For example, variants may be created using error prone PCR. In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Error prone PCR is described in Leung et al. (1989), *Technique* 1:11-15 and Caldwell et al. (1992), *PCR Methods Applic.* 2:28-33,

20 Briefly, in such procedures, nucleic acids to be mutagenized are mixed with PCR primers and reagents (e.g., reaction buffer, MgCl<sub>2</sub>, MnCl<sub>2</sub>, Taq polymerase and an appropriate concentration of dNTPs) for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction may be performed using 20 fmoles of nucleic acid to be mutagenized, 30 pmoles of each PCR  
 25 primer, a reaction buffer comprising 50mM KCl, 10mM Tris HCl (pH 8.3) and 0.01% gelatin, 7mM MgCl<sub>2</sub>, 0.5mM MnCl<sub>2</sub>, 5 units of Taq polymerase, 0.2mM dGTP, 0.2mM dATP, 1mM dCTP, and 1mM dTTP. PCR may be performed for 30 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min. However, it will be appreciated that these parameters may be varied as appropriate. The mutagenized nucleic acids are cloned into an  
 30 appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids are evaluated.

[000224] Variants also may be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis



is described in Reidhaar-Olson et al. (1988), *Science*, 241:53-57.

Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized. Clones containing the mutagenized DNA are recovered and the activities of the polypeptides they encode are assessed.

#### Assembly PCR

[000225] Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in U.S. Pat. No. 5,965,408.

#### Sexual PCR mutagenesis

[000226] Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different but highly related DNA sequence *in vitro*, as a result of random fragmentation of the DNA molecule based on sequence homology, followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described in Stemmer (1994), *Proc. Natl. Acad. Sci. USA* 91:10747-10751.

Briefly, in such procedures a plurality of nucleic acids to be recombined are digested with DNase to generate fragments having an average size of about 50-200 nucleotides. Fragments of the desired average size are purified and resuspended in a PCR mixture. PCR is conducted under conditions which facilitate recombination between the nucleic acid fragments. For example, PCR may be performed by resuspending the purified fragments at a concentration of 10-30 ng/μl in a solution of 0.2mM of each dNTP, 2.2mM MgCl<sub>2</sub>, 50mM KCl, 10mM Tris HCl, pH 9.0, and 0.1% Triton\* X-100. 2.5 units of Taq\* polymerase per 100μl of reaction mixture is added and PCR is performed using the following regime: 94°C for 60 seconds, 94°C for 30 seconds, 50-55°C for 30 seconds, 72°C for 30 seconds (30-45 times) and 72°C for 5 minutes. However, it will be appreciated that these parameters may be varied as appropriate. In some aspects, oligonucleotides may be included in the PCR reactions. In other aspects, the Klenow fragment of DNA polymerase I may be used in a first set of PCR reactions and Taq polymerase may be used in a subsequent

\*Trade-mark



set of PCR reactions. Recombinant sequences are isolated and the activities of the polypeptides they encode are assessed.

#### In vivo Mutagenesis

[000227] Variants may also be created by *in vivo* mutagenesis. In some aspects, random mutations in a sequence of interest are generated by propagating the sequence of interest in a bacterial strain, such as an *E. coli* strain, which carries mutations in one or more of the DNA repair pathways. Such “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for *in vivo* mutagenesis are described in PCT Publication No. WO 91/16427.

#### Cassette Mutagenesis

[000228] Variants may also be generated using cassette mutagenesis. In cassette mutagenesis a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

#### Recursive Ensemble Mutagenesis

[000229] Recursive ensemble mutagenesis may also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described in Arkin et al. (1992), *Proc. Natl. Acad. Sci. USA*, 89:7811-7815.

#### Exponential Ensemble Mutagenesis

[000230] In some aspects, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described in Delegrave et al. (1993), *Biotechnology Research* 11:1548-1552.



Random and site-directed mutagenesis

[000231] Random and site-directed mutagenesis is described in Arnold (1993), *Current Opinions in Biotechnology* 4:450-455.

5 Shuffling Procedures

[000232] In some aspects, the variants are created using shuffling procedures wherein portions of a plurality of nucleic acids which encode distinct polypeptides are fused together to create chimeric nucleic acid sequences which encode chimeric polypeptides as described in U.S. Patent. Nos. 5,965,408 and 5,939,250.

[000233] The variants of the polypeptides of the Group B amino acid sequences may be variants in which one or more of the amino acid residues of the polypeptides of the Group B amino acid sequences are substituted with a conserved or non-conserved amino acid residue (e.g, a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.

[000234] Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the following replacements: replacements of an aliphatic amino acid such as Alanine, Valine, Leucine and Isoleucine with another aliphatic amino acid; replacement of a Serine with a Threonine or vice versa; replacement of an acidic residue such as Aspartic acid and Glutamic acid with another acidic residue; replacement of a residue bearing an amide group, such as Asparagine and Glutamine, with another residue bearing an amide group; exchange of a basic residue such as Lysine and Arginine with another basic residue; and replacement of an aromatic residue such as Phenylalanine, Tyrosine with another aromatic residue.

[000235] Other variants are those in which one or more of the amino acid residues of the polypeptides of the Group B amino acid sequences includes a substituent group.

[000236] Still other variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

[000237] Additional variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretory sequence, a proprotein sequence or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.



[000238] In some aspects, the fragments, derivatives and analogs retain the same biological function or activity as the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto. In other aspects, the fragment, derivative, or analog includes a proprotein, such that the fragment, derivative, or analog can be activated by cleavage of the proprotein portion to produce an active polypeptide.

[000239] Another aspect of the invention is polypeptides or fragments thereof which have at least about 85%, at least about 90%, at least about 95%, or more than about 95% homology to one of the polypeptides of the Group B amino acid sequences, sequences substantially identical thereto, or a fragment comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. Percent identity may be determined using any of the programs described above which aligns the polypeptides or fragments being compared and determines the extent of amino acid homology or similarity between them. It will be appreciated that amino acid "homology" includes conservative amino acid substitutions such as those described above. In one aspect of the invention, the fragments can be used to generate antibodies. These antibodies can be used to immobilize nitrilases can be used in industrial processes. Polynucleotides encoding the nitrilases of the present invention can be used in a similar way.

[000240] Alternatively, the homologous polypeptides or fragments may be obtained through biochemical enrichment or purification procedures. The sequence of potentially homologous polypeptides or fragments may be determined by proteolytic digestion, gel electrophoresis and/or microsequencing. The sequence of the prospective homologous polypeptide or fragment can be compared to one of the polypeptides of the Group B amino acid sequences, sequences substantially identical thereto, or a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof using any of the programs described herein.

[000241] Another aspect of the invention is an assay for identifying fragments or variants of the Group B amino acid sequences, or sequences substantially identical thereto, which retain the enzymatic function of the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto. For example, the fragments or variants of the polypeptides, may be used to catalyze biochemical reactions, which indicate that said fragment or variant retains the enzymatic activity of the polypeptides in Group B amino acid sequences.



[000242] The assay for determining if fragments of variants retain the enzymatic activity of the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto includes the steps of: contacting the polypeptide fragment or variant with a substrate molecule under conditions which allow the polypeptide fragment or variant to function, and  
5 detecting either a decrease in the level of substrate or an increase in the level of the specific reaction product of the reaction between the polypeptide and substrate.

[000243] The polypeptides of the Group B amino acid sequences, sequences substantially identical thereto or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof may be used in a variety of applications. For example,  
10 the polypeptides or fragments thereof may be used to catalyze biochemical reactions. In accordance with one aspect of the invention, there is provided a process for utilizing a polypeptide of the Group B amino acid sequences, and sequences substantially identical thereto or polynucleotides encoding such polypeptides for hydrolyzing aminonitriles. In such procedures, a substance containing a haloalkane compound is contacted with one of the  
15 polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto under conditions which facilitate the hydrolysis of the compound.

[000244] Antibodies - The polypeptides of Group B amino acid sequences, sequences substantially identical thereto or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may also be used to generate antibodies  
20 which bind specifically to the enzyme polypeptides or fragments. The resulting antibodies may be used in immunoaffinity chromatography procedures to isolate or purify the polypeptide or to determine whether the polypeptide is present in a biological sample. In such procedures, a protein preparation, such as an extract, or a biological sample is contacted with an antibody capable of specifically binding to one of the polypeptides of the Group B  
25 amino acid sequences, sequences substantially identical thereto, or fragments of the foregoing sequences.

[000245] In immunoaffinity procedures, the antibody is attached to a solid support, such as a bead or column matrix. The protein preparation is placed in contact with the antibody under conditions under which the antibody specifically binds to one of the polypeptides of the  
30 Group B amino acid sequences, sequences substantially identical thereto, or fragments thereof. After a wash to remove non-specifically bound proteins, the specifically bound polypeptides are eluted.



[000246] The ability of proteins in a biological sample to bind to the antibody may be determined using any of a variety of procedures familiar to those skilled in the art. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. Particular assays include ELISA assays, sandwich assays, radioimmunoassays, and Western Blots.

[000247] The antibodies of the invention can be attached to solid supports and used to immobilize nitrilases of the present invention. Such immobilized nitrilases can be used, as described above, in industrial chemical processes for the conversion of nitriles to a wide range of useful products and intermediates.

[000248] Polyclonal antibodies generated against the polypeptides of the Group B amino acid sequences, and sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies which may bind to the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from cells expressing that polypeptide.

[000249] For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein (1975), *Nature*, 256:495-497, the trioma technique, the human B-cell hybridoma technique (Kozbor et al. (1983), *Immunology Today* 4:72, and the EBV-hybridoma technique (Cole et al. (1985), in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96,

[000250] Techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778 can be

adapted to produce single chain antibodies to the polypeptides of, for example, the Group B amino acid sequences, or fragments thereof. Alternatively, transgenic mice may be used to express humanized antibodies to these polypeptides or fragments.



[000251] Antibodies generated against a polypeptide of the Group B amino acid sequences, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof may be used in screening for similar polypeptides from other organisms and samples. In such techniques, polypeptides from the organism are contacted with the antibody and those polypeptides which specifically bind to the antibody are detected. Any of the procedures described above may be used to detect antibody binding. One such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in Enzymology*, 160:87-116.

#### 10            Use of Whole Cells Comprising A Nucleic Acid

[000252] The invention provides for the use of whole cells which have been transformed with nucleic acid (or an active fragment thereof) encoding one or more of the nitrilases of the invention. The invention also provides for the use of such a whole cell in performing a nitrilase reaction on a substrate. Therefore, this invention provides for methods of hydrolyzing a cyanohydrin or aminonitrile linkage using a whole cell comprising at least one nucleic acid or polypeptide disclosed herein (SEQ ID NOS:1-386). For example, a whole cell which is stably transfected (the invention also encompasses transiently transfected or transformed whole cells) with a nucleic acid encoding a nitrilase is one aspect of the invention. Such a cell is useful as a reagent in a reaction mixture to act on a substrate and exhibit nitrilase activity.

#### Sequence Analysis Software

[000253] Percent identity or homology between two or more sequences is typically measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI). Such software matches similar sequences by assigning a percent identity or homology to various deletions, substitutions and other modifications. The term "percent identity," in the context of two or more nucleic acids or polypeptide sequences, refers to the percentage of nucleotides or amino acid residues that are the same when compared after being aligned for maximum correspondence over a designated region or comparison "window." Under some algorithms, a conservative amino acid substitution can be considered "identical" and a change at a wobble site of a codon can be considered "identical."



[000254] "Alignment" refers to the process of lining up two or more sequences to achieve maximal correspondence for the purpose of assessing the degree of identity or homology, as defined within the context of the relevant alignment algorithm.

[000255] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated for a particular algorithm. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent identity or homology for the test sequences relative to the reference sequence, based on the program parameters.

[000256] A "comparison window", as used herein, is a segment of the contiguous positions in a nucleic acid or an amino acid sequence consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 nucleotides or residues, which may be compared to a reference sequence of the same or different number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (1981), *Adv. Appl. Math.* **2**:482, by the homology alignment algorithm of Needleman and Wunsch (1970), *J. Mol. Biol* **48**:443, by the search for similarity method of Pearson and Lipman (1988), *Proc. Natl. Acad. Sci. USA* **85**:2444-2448, by computerized implementations of these algorithms, or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, the BLAST program (Basic Local Alignment Search Tool, National Center for Biological Information), BESTFIT, FASTA, and TFASTA (Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Alignment of Multiple Protein Sequence), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLoCKs IMProved Searcher), Intervals and Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence



Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction and Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome  
 5 databases to identify polynucleotide sequences having substantially identical sequences. A number of genome databases are available, for example, a substantial portion of the human genome is available as part of the Human Genome Sequencing Project

(Gibbs, 1995). At

least twenty-one other genomes have already been sequenced, including, for example, *M.*  
 10 *genitalium* (Fraser et al., 1995), *M. jannaschii* (Bult et al., 1996), *H. influenzae* (Fleischmann et al., 1995), *E. coli* (Blattner et al., 1997), and yeast (*S. cerevisiae*) (Mewes et al., 1997), and *D. melanogaster* (Adams et al., 2000). Significant progress has also been made in sequencing the genomes of model organism, such as mouse, *C. elegans*, and *Arabidopsis sp.* Several databases containing genomic information annotated with some functional  
 15 information are maintained by different organizations, and are accessible via the internet,

[000257] Examples of useful algorithms are the BLAST and the BLAST 2.0 algorithms,  
 20 which are described in Altschul et al. (1977), *Nuc. Acids Res.* 25:3389-3402, and Altschul et al. (1990), *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

This algorithm involves first identifying high scoring  
 sequence pairs (HSPs) by identifying short words of length *W* in the query sequence, which  
 25 either match or satisfy some positive-valued threshold score *T* when aligned with a word of the same length in a database sequence. *T* is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased.  
 30 Cumulative scores are calculated using the parameter *M* (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity *X* from its maximum achieved value; the



cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. For nucleotide sequences, the BLASTN program uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989), *Proc. Natl. Acad. Sci. USA* 89:10915).

[000258] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993), *Proc. Natl. Acad. Sci. USA* 90:5873). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, less than about 0.01, or less than about 0.001.

[000259] In one aspect, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST"). In particular, five specific BLAST programs are used to perform the following task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and
- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

[000260] The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which may be obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are identified (*i.e.*, aligned) by



means of a scoring matrix, many of which are known in the art. In one example, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al. (1992), *Science* 256:1443-1445; Henikoff and Henikoff (1993), *Proteins* 17:49-61). In another example, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds. (1978), *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). BLAST programs are accessible through the U.S. National Library of Medicine.

[000261] The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some aspects, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

[000262] In a particular aspect, the invention provides a method for modifying small molecules, comprising contacting a polypeptide encoded by a polynucleotide described herein or enzymatically active fragments thereof with a small molecule to produce a modified small molecule. A library of modified small molecules is tested to determine if a modified small molecule is present within the library which exhibits a desired activity. A specific biocatalytic reaction which produces the modified small molecule of desired activity is identified by systematically eliminating each of the biocatalytic reactions used to produce a portion of the library, and then testing the small molecules produced in the portion of the library for the presence or absence of the modified small molecule with the desired activity.

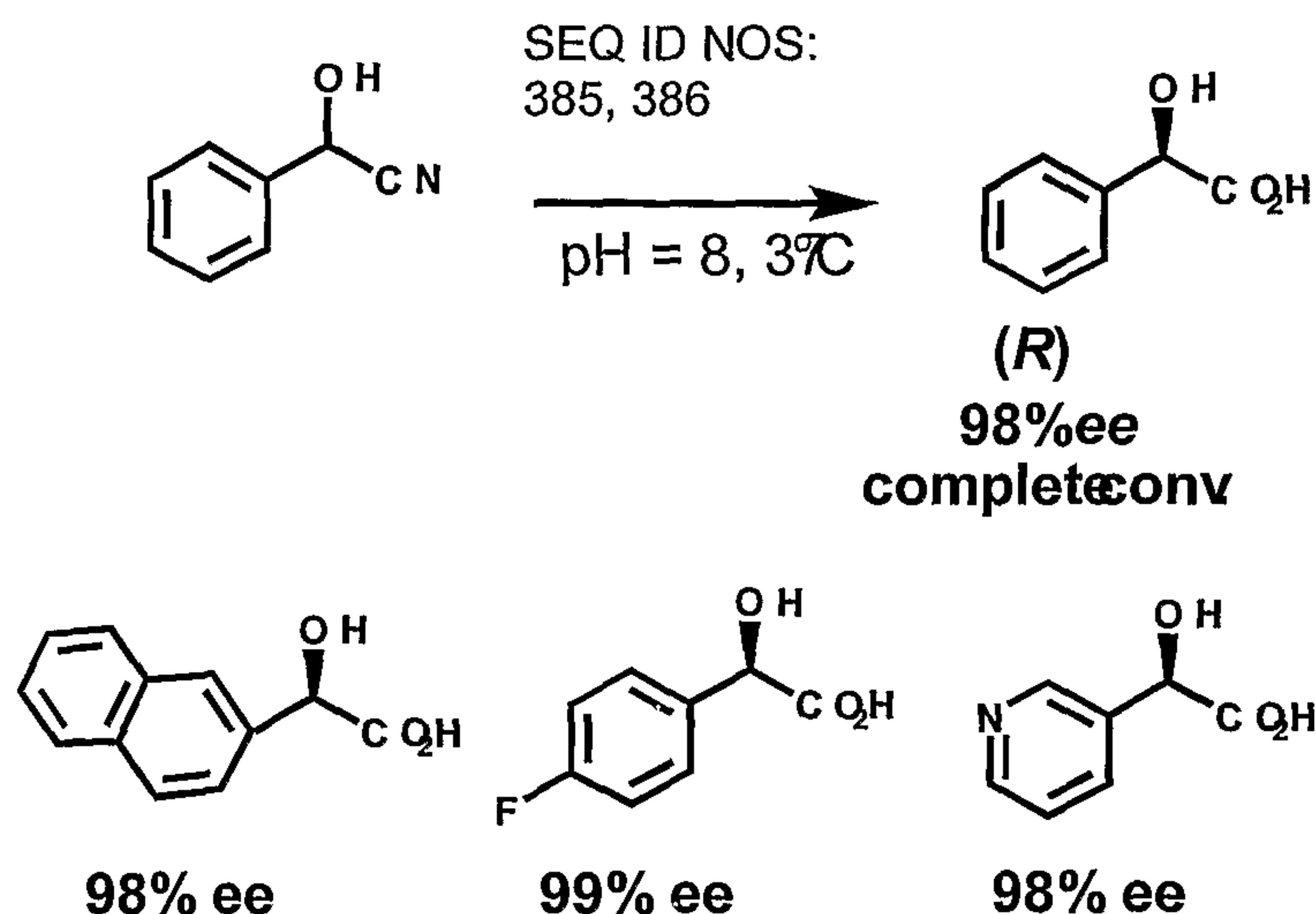
The specific biocatalytic reactions, which produce the modified small molecule of, desired activity is optionally repeated. The biocatalytic reactions are conducted with a group of biocatalysts that react with distinct structural moieties found within the structure of a small molecule, each biocatalyst is specific for one structural moiety or a group of related structural moieties; and each biocatalyst reacts with many different small molecules which contain the distinct structural moiety.

[000263] Some aspects of the use of the nitrilases are:

$\alpha$ -hydroxy acid - Nitrilases produce  $\alpha$ -hydroxy acids through hydrolysis of cyanohydrins. Production of mandelic acid and derivatives thereof is an example of this. A significant application of this type involves commercial production of (*R*)-mandelic acid in both high yield and high enantioselectivity from mandelonitrile. Mandelic acid and derivatives have found broad application as intermediates and resolving agents for the production of many chiral pharmaceutical and agricultural products. Previous attempts to

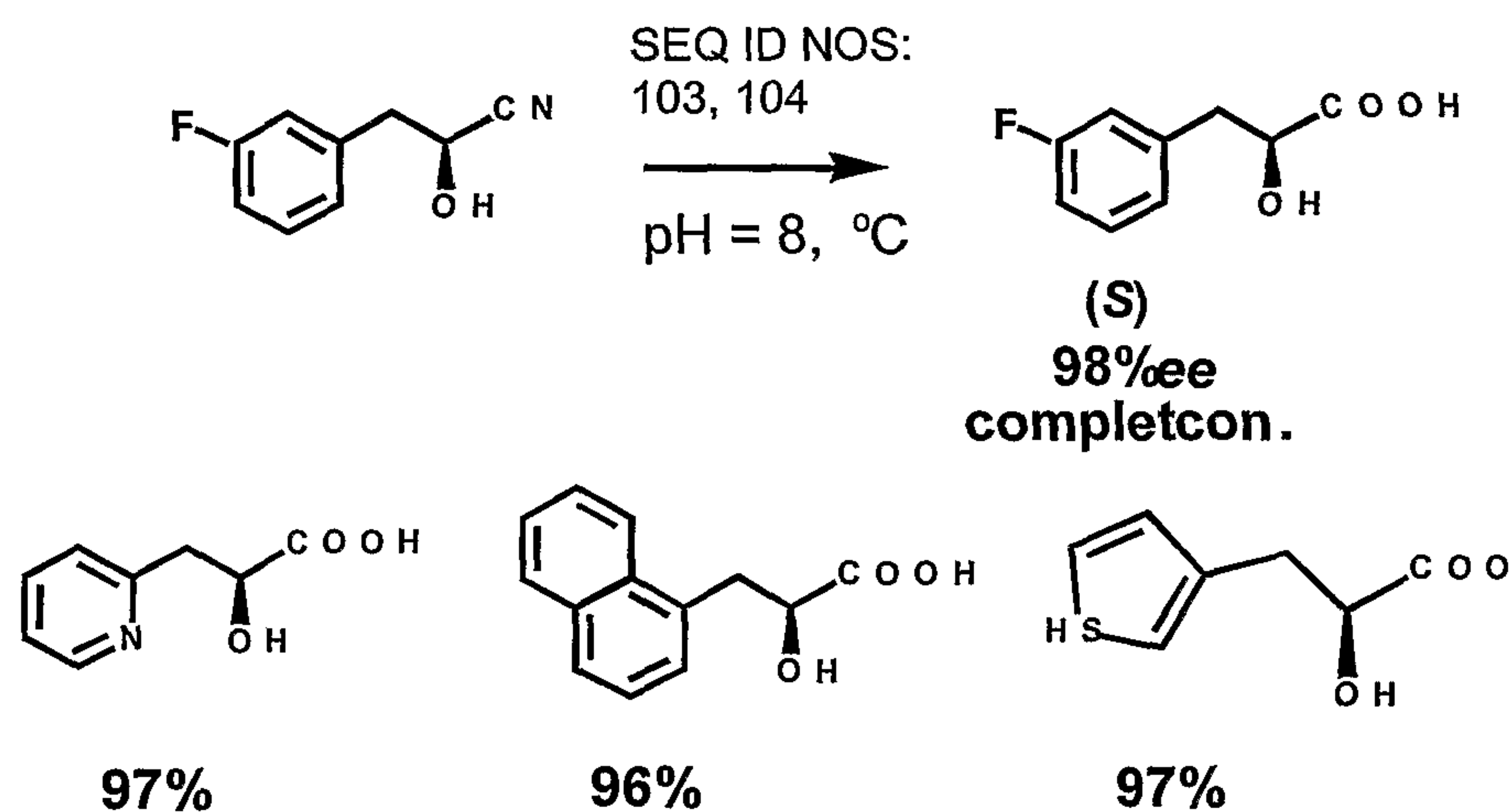


employ the few known nitrilases in processes using analogous substrates have been plagued by significantly lower activity, productivity, and selectivity.



#### Phenyllactic acid derivatives

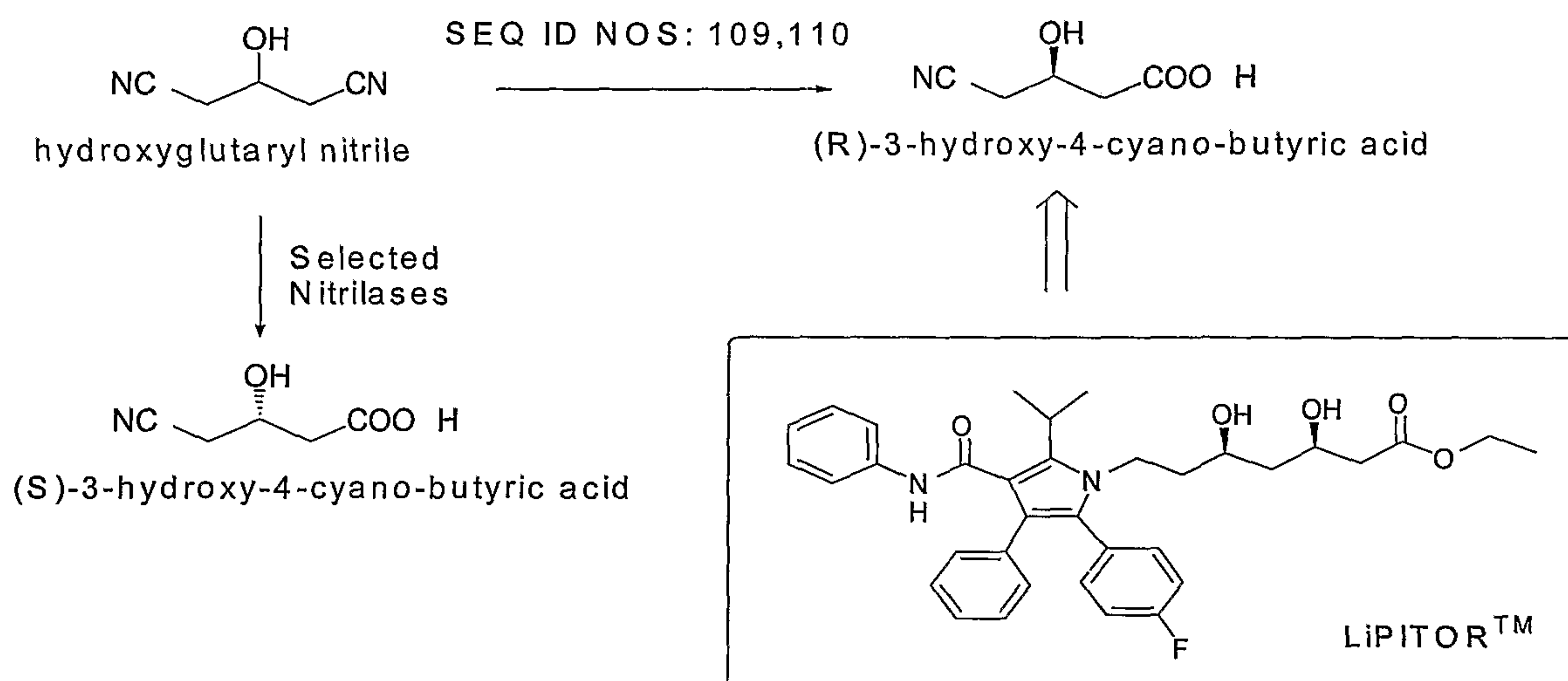
[000264] An additional application is in the production of (*S*)-phenyl lactic acid derivatives in both high yield and high enantioselectivity. Phenyl lactic acid derivatives have found broad application in the production of many chiral pharmaceutical and agricultural products.



#### β-hydroxy acid

[000265] With important commercial considerations, nitrilases are provided produce either enantiomer of 4-cyano-3-hydroxybutyric acid, the (*R*)-enantiomer of which is a key intermediate in the synthesis of the drug LIPITOR™.





[000266] The following nitrilases are more examples of nitrilases useful in converting hydroxyglutaryl nitrile to (R)-3-hydroxy-4-cyano-butyric acid: SEQ ID NOS:205, 206, SEQ ID NOS:207, 208, SEQ ID NOS:195, 196, SEQ ID NOS:43, 44, SEQ ID NOS:321, 322, and SEQ ID NOS:237, 238. The above schematic indicates that "selected nitrilases" can be used to convert hydroxyglutaryl nitrile to (S)-3-hydroxy-4-cyano-butyric acid: SEQ ID NOS:107, 108, SEQ ID NOS:109, 110, SEQ ID NOS:111, 112, SEQ ID NOS:127, 128, SEQ ID NOS:129, 130, SEQ ID NOS:133, 134, SEQ ID NOS:113, 114, SEQ ID NOS:145, 146, SEQ ID NOS:101, 102, SEQ ID NOS:179, 180, SEQ ID NOS:201, 202, SEQ ID NOS:159, 160, SEQ ID NOS:177, 178, SEQ ID NOS:181, 182, SEQ ID NOS:183, 184, SEQ ID NOS:185, 186, SEQ ID NOS:57, 58, SEQ ID NOS:197, 198, SEQ ID NOS:59, 60, SEQ ID NOS:67, 68, and SEQ ID NOS:359, 360.

[000267] The invention will be further described with reference to the following examples; however, it is to be understood that the invention is not limited to such examples. Rather, in view of the present disclosure which describes the current best mode for practicing the invention, many modifications and variations would present themselves to those of skill in the art without departing from the scope and spirit of this invention. All changes, modifications, and variations coming within the meaning and range of equivalency of the claims are to be considered within their scope.

## EXAMPLES



Example 1: Phagemid infections

[000268] For each library to be screened for nitrilases, an infection was set up as follows: 5ml of an  $OD_{600nm}=1$  resuspension of SEL700 cells and 1ml of the phagemid library to be screened were combined. The combination was incubated in a 37°C waterbath for 45 min.

- 5 [000269] Using the infection, serial dilutions were made in 10mM  $MgSO_4$ , using 10 $\mu$ l aliquots of the infection.

|    |                         |                                            |
|----|-------------------------|--------------------------------------------|
|    | <u>titer of library</u> | <u>dilutions to make</u>                   |
|    | $\sim 10^5$ cfu/ml      | $10^{-1}$ dilution                         |
|    | $\sim 10^6$ cfu/ml      | $10^{-1}$ , $10^{-2}$ dilution             |
| 10 | $\sim 10^7$ cfu/ml      | $10^{-1}$ , $10^{-2}$ , $10^{-3}$ dilution |

[000270] 60 $\mu$ l of each of the following dilutions were deposited onto a small LB-kan<sup>50</sup> plate:

|    |                         |                                         |
|----|-------------------------|-----------------------------------------|
|    | <u>titer of library</u> | <u>dilutions to make</u>                |
|    | $\sim 10^5$ cfu/ml      | undiluted infection, $10^{-1}$ dilution |
|    | $\sim 10^6$ cfu/ml      | $10^{-1}$ , $10^{-2}$ dilutions         |
| 15 | $\sim 10^7$ cfu/ml      | $10^{-2}$ , $10^{-3}$ dilutions         |

- [000271] The cells in the infection were centrifuged in a tabletop centrifuge at 4°C, 4.6k rpm, 10 min to form pellets. The supernatant was decanted from the resulting pellets. The cells were resuspended in residual liquid. All of the resuspended cells were deposited onto a single large LB-kan<sup>50</sup> plate. All plates were incubated at 30°C overnight.
- 20

Example 2: Selection Screenings

- [000272] The cells of each infection plate were resuspended with ~4mls 10mM  $MgSO_4$ . The resuspensions were placed in a tube. The remaining cells on each plate were resuspended with ~3mls 10mM  $MgSO_4$  and combined with the first resuspension from the same plate. The volume of each tube was brought to 12ml with 10mM  $MgSO_4$ . The tubes were vortexed vigorously. The tubes were centrifuged in a tabletop centrifuge at 4°C and 4.6k for 10min to form pellets. The supernatant was decanted from each resuspension. The washed cells in each tube were resuspended with 10ml 10mM  $MgSO_4$ . The resuspensions from each library were stored at 4°C until the selection cultures were ready to be set up.
- 25
- 30

[000273] For each resuspension, selection cultures were set up using the following process:



- 1) The nitrilase selection medium was prepared, using: 1XM9 medium with 0.2% glucose, no nitrogen and 50µg/ml kanamycin (for pBK phagemid libraries only; use ampicillin for pBS libraries).
- 2) 5ml of the medium was aliquoted into a 50ml screw top conical tube.
- 3) 25µl of the stored resuspension was added to the tube.
- 4) 5µl of adiponitrile was added to the tube, to bring the final concentration to 8.8mM. Additional nitrile substrates may be used, in place of adiponitrile.
- 5) The resulting combination was cultured at 30°C.

Steps 1-5 were repeated for each nitrile substrate.

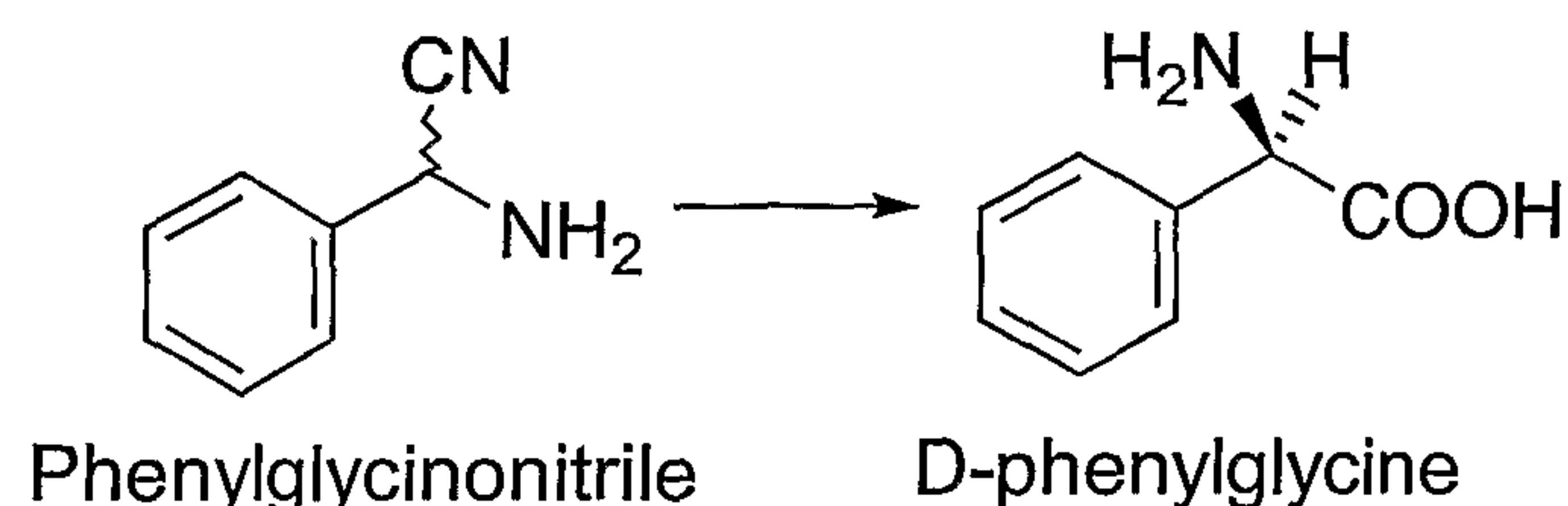
#### Example 3: Isolation of a positive nitrilase clone from selection cultures

[000274] Ten (10) µl of selection culture with growth was streaked out onto a small LB-kan<sup>50</sup> plate and allowed to grow for 2 nights at 30°C. Five isolated cfu were picked and each was grown in 2ml nitrilase selection medium at 30°C. Each culture was monitored (where growth indicates positive cfu was picked), and was removed when monitoring indicated that it was in a stationary phase of growth. One (1) ml of culture was used to do a plasmid preparation and was eluted with 40µl elution buffer. Five to eight (5-8) µl DNA was cut with Pst I/Xho I or Sac I/Kpn I restriction enzymes to remove insert from vector. A restriction fragment length polymorphism (RFLP) determination was carried out to identify the size of the insert. The insert was sequenced.

#### Example 4: Screening and Characterization of Nitrilases

[000275] Nitrilases of the invention were screened against target substrates. Of those showing hydrolytic activity in a primary screen, enzymes with enantioselectivities above 20% enantiomeric excess (*ee*) were selected for further characterization. Those enzymes were selected based on: 1) having activity against one of the substrates of interest and 2) exhibition of greater than 35% *ee* (enantiomeric excess). The results of this screening process are set forth in Table 1 above. The products used for screening were: D-Phenylglycine, L-Phenyllactic acid, (*R*) 2-chloromandelic acid, (*S*)- Cyclohexylmandelic acid, L-2-methylphenylglycine, (*S*)-2-amino-6-hydroxy hexanoic acid, and 4-methyl-L-leucine.



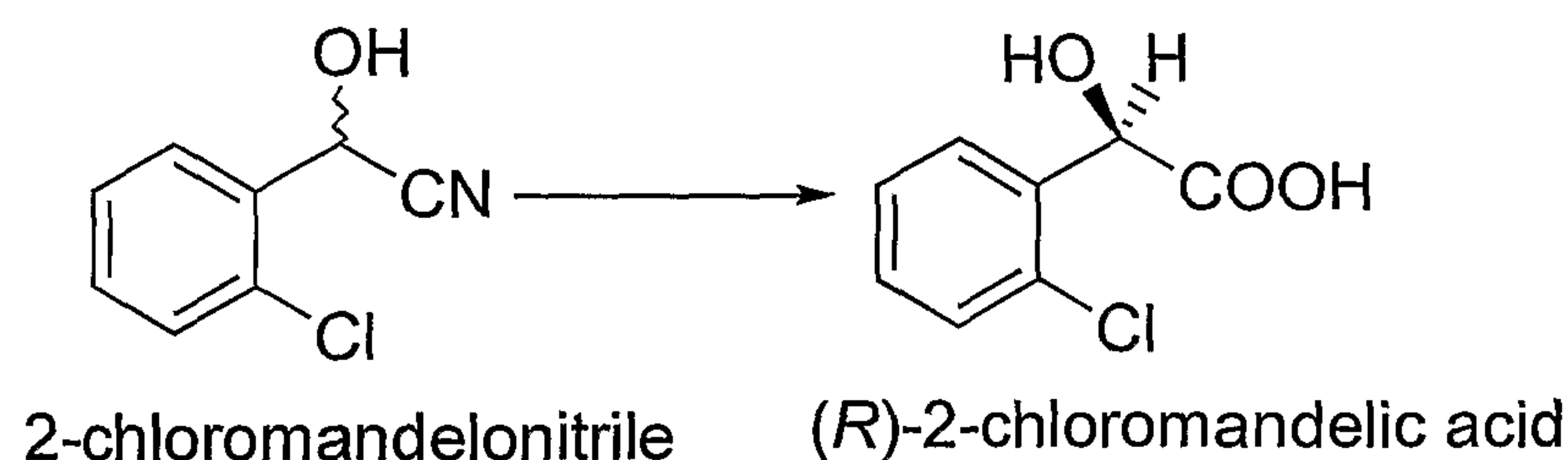
Screening of nitrilases against target substrate D-Phenylglycine

[000276] The hydrolysis of phenylglycinonitrile was performed. Some of these enzymes  
5 showed an *ee* higher than 20% and those were selected for preliminary characterization.

[000277] Based on the preliminary characterization experiments, a number of putative hits  
were identified on phenylglycinonitrile and a large amount of data was accumulated on these  
enzymes. The data revealed many common properties: the majority of the enzymes had pH  
optima for activity at pH 7 and, in general, the enantioselectivity was enhanced at the lower  
10 pH values. The enzymes were found to be more active at higher temperature, particularly  
38°C, although this temperature often resulted in lower enantioselectivities. The use of  
water-miscible co-solvents in the reaction was shown to be a practical option. The inclusion  
of 10-25% methanol (v/v) in the enzyme reactions did not substantially affect enzyme  
activity and in many cases, led to an increase in enantioselectivity. The use of biphasic  
15 systems has also shown some promise, with the enzymes maintaining their level of activity  
with the addition of up to 70% (v/v) of hexane and, in some cases, toluene. The use of ethyl  
acetate in the biphasic systems, however, led to lower activity.

[000278] Of the enzymes identified active on phenylglycinonitrile, the enantioselectivity of  
several enzymes was shown to remain above the success criterion of 35% *ee*. The  
20 preliminary characterization data indicated that some of the enzymes exhibited high  
enantioselectivities for D-phenylglycine, with corresponding conversion to product of 40-  
60%. Further investigation suggested that the rate of activity of some of these enzymes was  
faster than the rate of racemization of the substrate. Reducing the concentration of enzyme  
led to improved enantioselectivity; therefore, it appears that some benefit could be gained by  
25 control of the relative rates of the chemical racemization and the enzyme activity.



Screening of nitrilases against target substrate (*R*)-2-chloromandelic acid

[000279] Enzymes were identified which showed activity on 2-chloromandelonitrile. A high degree of overlap existed between the enzymes which were active on 2-chloromandelonitrile and phenylglycinonitrile. Many of these enzymes also formed a distinct sequence family.

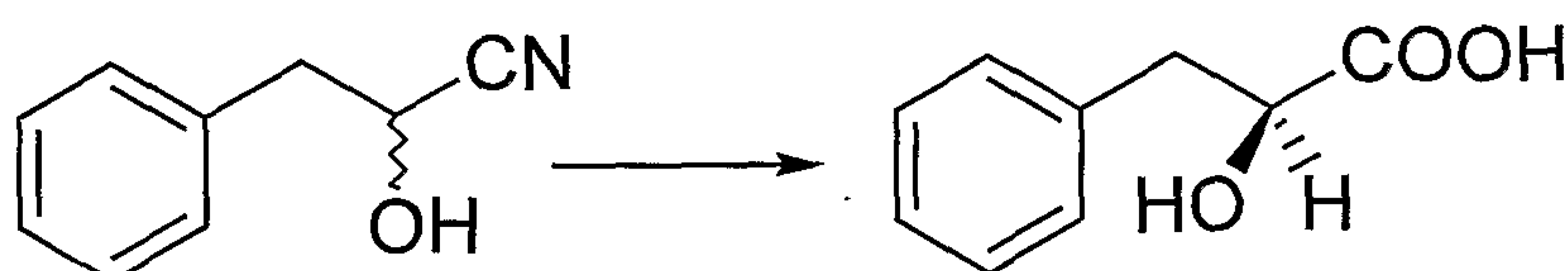
[000280] Higher temperatures and neutral pH appeared to lead to the highest activity for the active enzymes. For the majority of the nitrilases, the enantioselectivity also increased at higher temperatures, particularly 38°C. The enzymes retained their activity in the presence of up to 25% methanol or 10% isopropanol; in many of these cases, the enantioselectivity was also enhanced. Activity in biphasic systems was largely comparable to aqueous conditions, particularly with hexane as the non-aqueous phase; varying tolerances to toluene were observed between the different nitrilases.

Table 2. Summary of optimal conditions determined from characterization experiments for enantioselective hydrolysis of 2-chloromandelonitrile.



| SEQ ID NOS: | Optimum pH | Optimum Temp °C | Solvent Tolerance                          |
|-------------|------------|-----------------|--------------------------------------------|
| 385, 386    | 7          | 38              | 25% MeOH                                   |
| 169, 170    | 5          | 38              | 25% MeOH, 10% IPA                          |
| 185, 186    | 7          | 38              | 25% MeOH, 10% IPA                          |
| 47, 48      | 7          | 38              | 10% MeOH                                   |
| 197, 198    | 6          | 55              | 25% MeOH, 10% IPA                          |
| 187, 188    | 7          | 38              | 10% MeOH; 40% IPA                          |
| 217, 218    | 7          | 38              | 25% MeOH, 10% IPA, 70% hexane, 40% toluene |
| 55, 56      | 7          | 38              | 10% MeOH, IPA, 70% hexane                  |
| 167, 168    | 9          | 38              | 10% MeOH, IPA, 70% hexane                  |
| 15, 16      | 7          | 38              | 25% MeOH, 10% IPA, 70% hexane, 40% toluene |

Screening of nitrilases against target substrate (*S*)-phenyllactic acid:



Phenylacetaldehyde  
cyanohydrin

(*S*)-Phenyllactic acid

5

[000281] Many of the nitrilases tested were active on phenylacetaldehyde cyanohydrin. Many of these enzymes were part of two related sequence families and were distinct from those enzymes that were active on phenylglycinonitrile and chloromandelonitrile.

[000282] The pH optima of the enzymes was generally above pH 7 (*i.e.* pH 8 or 9), with higher enantioselectivities being exhibited at these levels. Most of the enzymes showed superior activity at higher temperature, particularly 38°C. The effect of temperature on the enantioselectivities of the enzymes varied; in most cases, this property was slightly lower at

10



WO 03/097810

PCT/US03/15712

higher temperatures. While the enzymes were tolerant towards the addition of co-solvents, particularly 10% (v/v) methanol, no advantage in activity or enantioselectivity was gained by such additions. The use of a biphasic system was again shown to be feasible.

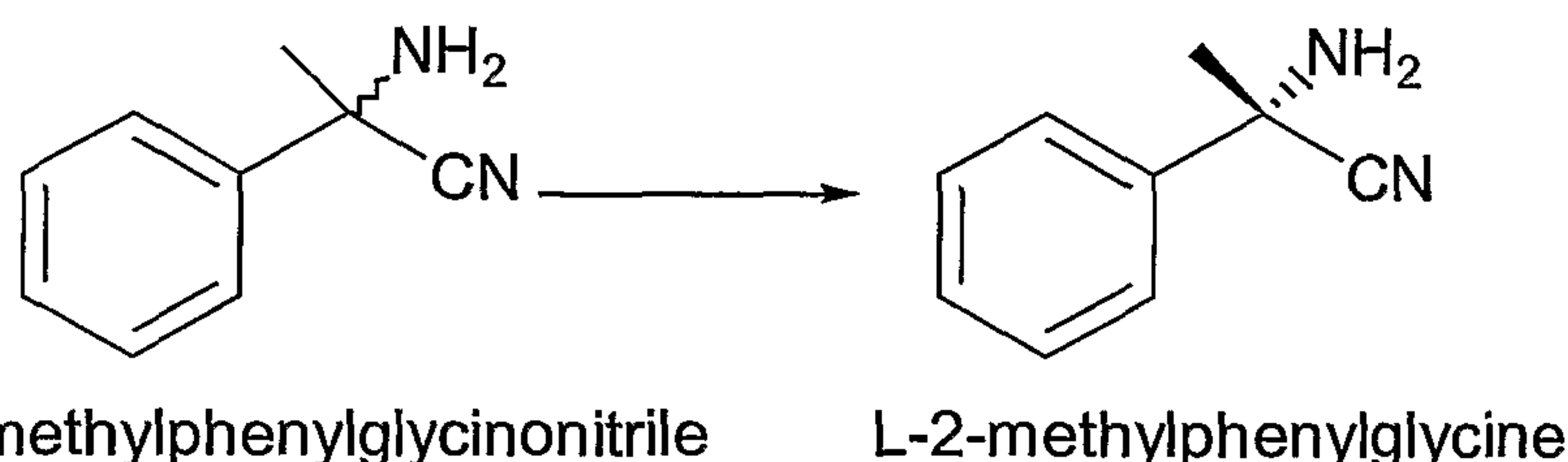
[000283] Table 3. Summary of optimal conditions determined from characterization experiments for enantioselective hydrolysis of phenylacetaldehyde cyanohydrin

| SEQ ID NOS: | Optimum pH | Optimum Temp °C | Solvent Tolerance                      |
|-------------|------------|-----------------|----------------------------------------|
| 103, 104    | 7          | 55              | 10% MeOH, IPA                          |
| 99, 100     | 8          | 38              | 10% MeOH, 70% hexane, toluene          |
| 183, 184    | 9          | 38              | 10% MeOH, IPA, 70% toluene, hexane     |
| 173, 174    | 5          | 38-55           | 25% MeOH, IPA, 70% hexane, toluene     |
| 213, 214    | 7          | 38              | 10% MeOH, 25% IPA, 70% hexane, toluene |
| 61, 62      | 7          | 38              | 10% MeOH, 70% hexane, toluene          |
| 205, 206    | 8          | 38-55           | 10% MeOH, IPA, 40% hexane, toluene     |
| 207, 208    | 8          | 38              | 10% MeOH, 70% hexane                   |
| 209, 210    | 8          | 38              | 10% MeOH, 40% hexane, toluene          |
| 195, 196    | 8          | 38              | 10% MeOH, 40% hexane, toluene          |
| 43, 44      | 9          | 38              | 10% MeOH, 40% hexane                   |
| 161, 162    | 9          | 38              | 25% MeOH, IPA, 10% hexane, toluene     |
| 175, 176    | 6          | 38-55           | 10% MeOH, IPA, 40% hexane              |



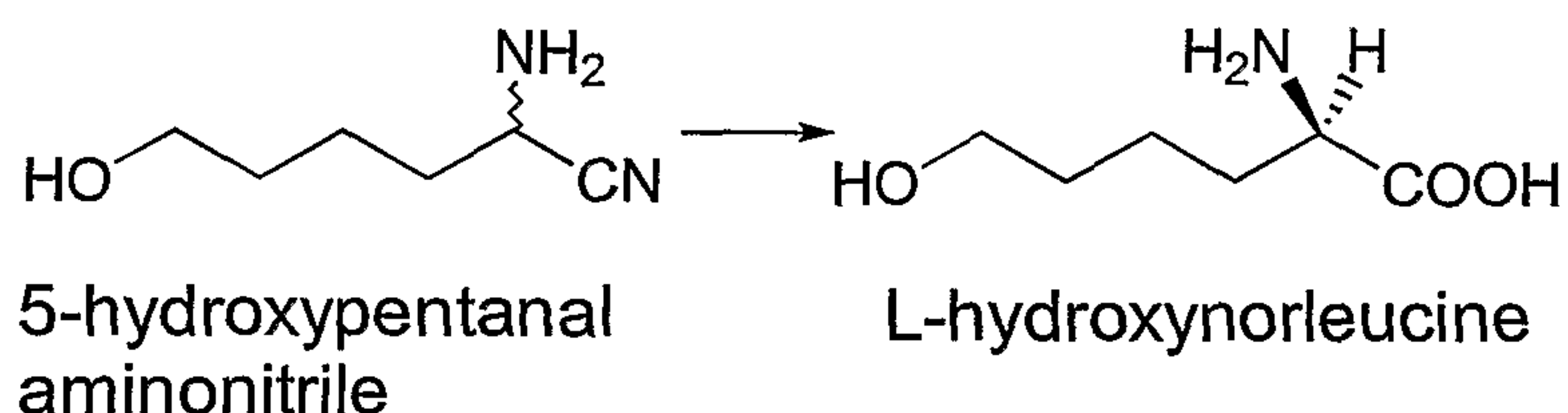
| SEQ ID NOS: | Optimum pH | Optimum Temp °C | Solvent Tolerance            |
|-------------|------------|-----------------|------------------------------|
| 293, 294    | 6          | 38              | 10% MeOH, IPA,<br>40% hexane |

Screening of nitrilases against target substrate L-2-methylphenylglycine



[000284] Nitrilases have shown activity on this substrate and preferentially yielded the D-2-methylphenylglycine, rather than the required L-2-methylphenylglycine.

Screening of nitrilases against target substrate L-hydroxynorleucine ((S)-2-amino-6-hydroxy hexanoic acid)



[000285] A number of nitrilases, which showed activity on 2-amino-6-hydroxy hexanenitrile, were isolated. All of these enzymes showed enantioselectivity towards the L-isomer of the product.

[000286] The enzymes all showed higher enantioselectivities at higher pH and appeared to more susceptible to the addition of solvents than the other nitrilases tested. Although activity was detected in the presence of organic solvents, it was generally lower than that of the aqueous control. Once again, the activity of the enzymes was negatively affected by the acid product and aldehyde starting material.

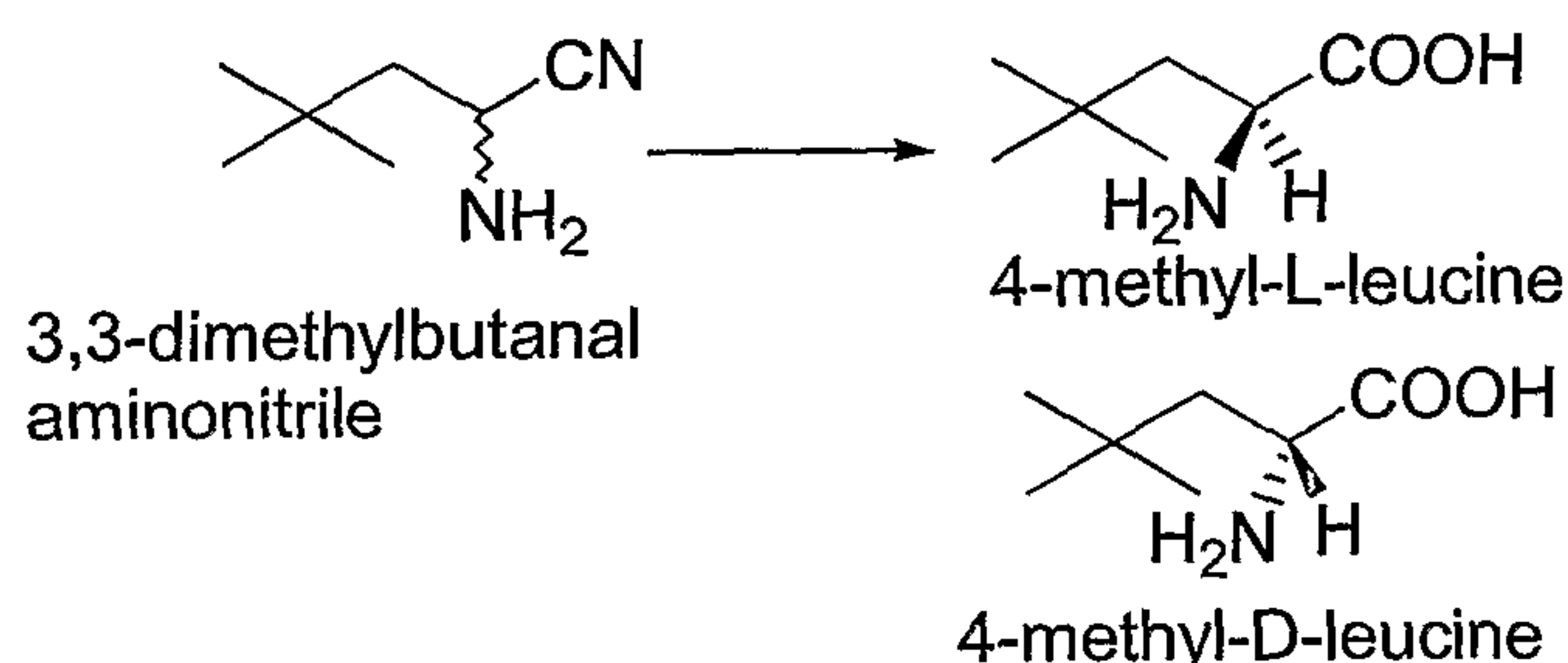
Table 4. Summary of optimal conditions determined from characterization experiments for enantioselective hydrolysis of 2-amino-6-hydroxy hexanenitrile.



| SEQ ID NOS: | Optimum pH | Optimum Temp °C | Solvent  |
|-------------|------------|-----------------|----------|
| 217, 218    | 9          | 38              | 10% MeOH |
| 55, 56      | 9          | 38              | None     |
| 187, 188    | 9          | 38              | 10% MeOH |
| 167, 168    | 9          | 38              | None     |
| 221, 222    | 9          | 38              |          |

[000287] A range of hydrolytic activities was observed among the confirmed hit enzymes for 2-amino-6-hydroxy hexanenitrile.

5 Screening of nitrilases against target substrate 4-methyl-D-leucine and 4-methyl-L-leucine

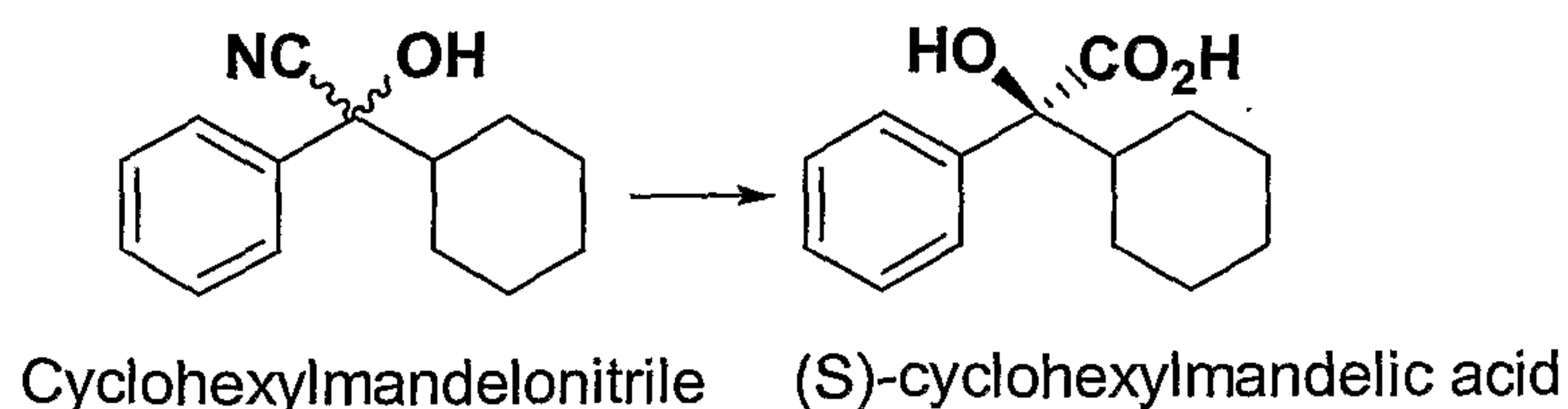


[000288] Hydrolysis of 2-amino-4,4-dimethyl pentanenitrile was performed by several of the nitrilases. Of these, some were shown to hydrolyse the nitrile to the L-isomer of the corresponding acid and were selected for further characterization.

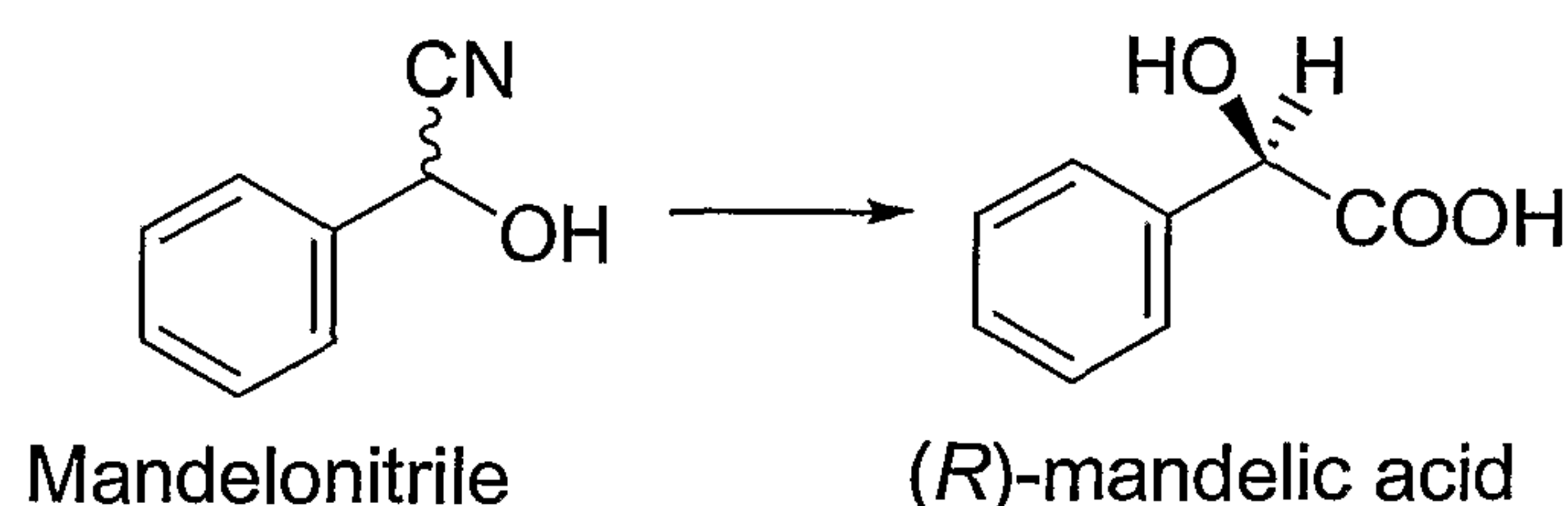
10 Table 5. Summary of optimal conditions determined from characterization experiments for enantioselective hydrolysis of 2-amino-4,4-dimethyl pentanenitrile

| SEQ ID NOS: | Optimum pH | Optimum Temp °C | Solvent Tolerance |
|-------------|------------|-----------------|-------------------|
| 103, 104    | 7          | 23              | 25% MeOH, 10% IPA |
| 59, 60      | 8          | 23              | 25% MeOH          |
| 221, 222    | 6          | 38              | 25% MeOH, 10% IPA |

Screening of nitrilases against target substrate (S)-cyclohexylmandelic acid





Screening of nitrilases against target substrate Mandelonitrile

- 5 [000289] The nitrilase collection was also screened on mandelonitrile. The nitrilases actively hydrolyzed both phenylglycinonitrile and chloromandelonitrile.

Enzymatic assay for determination of enantioselectivity

- [000290] In the design of a spectroscopic system for determination of the chiral  $\alpha$ -hydroxy acids and  $\alpha$ -amino acids, an enzyme based assay which permits the detection of product  
10 formation and enantioselectivity was developed and used.

- [000291] Spectroscopic systems for the detection of  $\alpha$ -hydroxy- and for  $\alpha$ -amino- acids based on lactate dehydrogenase (L-LDH & D-LDH) and on amino acid oxidase (L-AA Oxid & D-AA Oxid) are described in Figures 6 and 7. These enzymes were chosen because they are reported to have reasonably broad substrate ranges while still retaining near absolute  
15 enantiospecificity.

- [000292] The overall feasibility of this system has been established (Table 12). Neither the parent hydroxynitrile nor the aminonitrile is metabolized by the secondary or detection enzyme and thus starting material does not interfere. Cell lysate which is not heat treated results in background activity for the LDH system; however, heat inactivation eliminates the  
20 background activity. Cell lysate does not appear to interfere in the AA Oxidase assay. One concern is the inactivation of the AA Oxidase, which utilizes a FMN co-factor, by residual cyanide. However, the control studies indicated that at 2 mM PGN (which could release up to 2 mM HCN) inactivation is not a problem. This assay is suitable for automation of 384 well (or possibly greater density) microtiter plates.

- 25 Table 6: Summary of Identification of Secondary Enzyme to Chiral Detection of Acid Product.



| SUBSTRATE                                | ENZYME WITH SUITABLE<br>ACTIVITY FOUND FROM<br>COMMERCIAL SOURCE |
|------------------------------------------|------------------------------------------------------------------|
| Hydroxy Acid Products:                   |                                                                  |
| L-lactic acid                            | YES                                                              |
| D-lactic acid                            | YES                                                              |
| L-phenyl lactic acid                     | YES                                                              |
| D-phenyl lactic acid                     | YES                                                              |
| S-cyclohexylmandelic acid <sup>1</sup>   | <i>Not applicable</i>                                            |
| R-cyclohexylmandelic acid <sup>1</sup>   | <i>Not applicable</i>                                            |
| Amino Acid Products:                     |                                                                  |
| 4-methyl-L-leucine                       | YES                                                              |
| 4-methyl-L/D-leucine                     | YES (D-unknown)                                                  |
| D-phenylalanine                          | YES                                                              |
| R-phenylglycine                          | YES                                                              |
| L-homophenyllactic acid                  | YES                                                              |
| D-homophenyllactic acid                  | YES                                                              |
| L-homophenylalanine                      | YES                                                              |
| D-homophenylalanine                      | YES                                                              |
| (S)-2-amino-6-hydroxy<br>hexanoic acid   | YES                                                              |
| (R/S)-2-amino-6-hydroxy<br>hexanoic acid | YES (D-unknown)                                                  |
| L-methylphenylglycine <sup>1</sup>       | 1. <i>Not Applicable</i>                                         |
| D-methylphenylglycine <sup>1</sup>       | <i>Not Applicable</i>                                            |

<sup>1</sup>: The assay will not be applicable to cyclohexylmandelic acid and 2-methylphenylglycine, as tertiary alcohols are not amenable to this particular oxidation

#### 5 Example 5: Standard assay conditions

[000293] The following solutions were prepared:

- Substrate stock solution: 50 mM of the aminonitrile substrate in 0.1 M phosphate buffer (pH 7) or 50 mM of the cyanohydrin substrate in 0.1 M Na Acetate buffer (pH 5)
- Enzyme stock solution: 3.33 ml of 0.1 M phosphate buffer (pH 7) to each vial of 20 mg of lyophilized cell lysate (final concentration 6 mg protein/ml)

[000294] Procedure:

- Add 100  $\mu$ l of the 50 mM substrate solution to the appropriate number of wells of a 96-well plate
- Add 80  $\mu$ l of buffer to each well
- Add 20  $\mu$ l of enzyme solution to each well
- Blank controls were set up by substitution of 20  $\mu$ l of buffer for the enzyme solution



• Negative controls consisting of 20 µl of enzyme solution in 180 µl of buffer were also included in many of the experiments. Once it had been established that the cell lysate did not interfere with the detection of the products, these controls were not included.

[000295] Sampling of reactions:

5       • The reactions were sampled by removing an aliquot from each well (15-50 µl) and diluting the samples as follows:

• Samples for non-chiral HPLC analysis:

10       • Phenylglycine, 2-chloromandelic acid and phenyllactic acid: initially, the samples were diluted 2-fold with water and a further 2-fold with methanol or acetonitrile (final dilution: 4-fold). It was found that an 8-fold dilution of these samples led to improved chromatographic separation

15       • (S)-2-amino-6-hydroxy hexanoic acid, 4-methylleucine, *t*-leucine, 2-methylphenylglycine and cyclohexylmandelic acid: samples were diluted 1:1 with methanol or acetonitrile. The choice of solvent was based on the solvent used in the HPLC analysis method.

[000296] • Samples for chiral HPLC analysis:

• Phenylglycine, 2-chloromandelic acid and phenyllactic acid: as described above for the non-chiral analyses, the samples for chiral analyses were initially diluted 2-fold and in the later stages of the project, at 4-fold.

20       • (S)-2-amino-6-hydroxy hexanoic acid, 4-methylleucine, *t*-leucine, 2-methylphenylglycine: samples were diluted 1:1 with methanol or acetonitrile.

[000297] • For each experiment, a standard curve of the product was included in the HPLC run. The curve was plotted on an X-Y axis and the concentration of product in the samples calculated from the slope of these curves.

25 [000298] • For the preliminary characterization experiments, samples were taken such that the activity of the enzymes was in the linear phase; this was performed so that differences in the effects of the parameters on the rate of reaction, rather than the complete conversion, could be determined. The sampling times are denoted in the tables included in the text.

[000299] • The samples were analyzed by HPLC, using the methods outlined in Table 20 and 21.



Example 6: Determination of the Effect of pH on enzyme activity and enantioselectivity

[000300] The effect of pH on the enzyme activity and enantioselectivity was studied by performance of the standard assay in a range of different buffers:

0.1 M Citrate Phosphate pH 5

5 0.1 M Citrate Phosphate pH 6

0.1 M Sodium Phosphate pH 7

0.1 M Tris-HCl pH 8

M Tris-HCl pH 9

10 [000301] The samples were analyzed by non-chiral and chiral HPLC methods and examples of the results are presented in Tables 5, 8 and 11 herein.

Example 7: Determination of the effect of temperature on enzyme activity and enantioselectivity

15 [000302] The effect of temperature on the activity and enantioselectivity was investigated by performing the standard assay at room temperature, 38°C and 55°C. The samples were analyzed by non-chiral and chiral HPLC methods and examples of the results are given in Tables 5, 8 and 11 herein.

Example 8: Determination of the Effect of solvents on enzyme activity and enantioselectivity

20 [000303] The enzyme reactions were performed in the presence of cosolvents and as biphasic systems, in order to investigate the effect of water-miscible and water-immiscible solvents on the enzymes. In the presence of cosolvents, the reactions were run under standard conditions, with substitution of the buffer with methanol or isopropanol. The final concentrations of solvent in the reactions was 0, 10, 25 and 40% (v/v).

25 [000304] The biphasic reactions were also carried out under standard conditions, with a layer of water-immiscible organic solvent forming the nonaqueous phase. The solvent was added at the following levels: 0%, 10%, 40% and 70% (v/v) of the aqueous phase. The samples from these reactions were evaporated by centrifugation under vacuum and redissolved in a 50:50 mixture of methanol or acetonitrile and water. The samples were  
30 analyzed by non-chiral and chiral HPLC methods.



Example 9: Determination of the Effect of process components on enzyme activity and enantioselectivity

Activity

[000305] The effect of the process components on the activity of the enzymes was established by addition of the individual components to the enzymatic reaction. These components included the starting materials for the nitrile synthesis, aldehyde, cyanide and ammonium, as well as triethylamine, which is added in catalytic amounts to the nitrile synthesis reaction. The concentrations of the reactants were selected with possible process conditions in mind and were adapted to the levels of reactants used in the enzyme assays. In some cases, the solubility of the aldehydes and products was relatively low; in these cases, the highest level of solubility was added to the reactions as the highest level and 10% of this level as the lower value.

[000306] The enzymatic reactions were carried out under standard conditions, with addition of one or more of the following components: benzaldehyde, phenylglycine, phenylacetaldehyde, phenyllactic acid, 2-chlorobenzaldehyde, 2-chloromandelic acid, 5-hydroxypentanal, (S)-2-amino-6-hydroxy hexanoic acid, 4-methylleucine, KCN, Triethylamine, NH<sub>4</sub>Cl. Control reactions were performed under standard conditions, with no additive. The samples were analyzed by non-chiral HPLC.

Stability

[000307] The stability of the enzymes to process conditions was monitored by incubation of the enzymes in the presence of the individual reaction components for predetermined time periods, prior to assay of the enzyme activity under standard conditions. In these experiments, the enzymes were incubated at a concentration of 1.2 mg protein/ml in the presence of each of the following reaction components: methanol, benzaldehyde, phenylglycine, phenylacetaldehyde, phenyllactic acid, 2-chlorobenzaldehyde, 2-chloromandelic acid, 5-hydroxypentanal, (S)-2-amino-6-hydroxy hexanoic acid, KCN, NH<sub>4</sub>Cl.

Assay conditions:

[000308] At 0, 2, 6 and 24 hours of incubation in the particular additive, 50 µl of the enzyme solution was removed, 50 µl of a 50 mM substrate stock solution added and the enzyme activity assayed under standard conditions. After substrate addition, the reactions were sampled at the following times: Phenylglycinonitrile: 10 mins; Phenylacetaldehyde cyanohydrin: 1 hour; 2-chloromandelonitrile: 2 hours. Control reactions were performed by



incubation of the enzyme in buffer only. The samples were analyzed using non-chiral HPLC methods.

Example 10: Confirmation of putative hit enzymes

5 [000309] Following the preliminary characterization experiments, the enzymes which were identified as putative hits were assayed under the optimal conditions determined, in order to evaluate their performance, especially in terms of enantioselectivity, when higher conversions were attained. The enzymes were assayed with 25 mM substrate, under the conditions of pH and temperature noted in the tables included in the text. A standard concentration of 0.6 mg/ml protein was used for each of the enzymes, unless otherwise stated.

10 Example 11: Selected examples of chromatograms from enzyme reactions

[000310] In this section, representative examples of chromatograms for each substrate and product combination will be shown, together with a discussion of some of the challenges encountered with the methods and how they were addressed.

D-Phenylglycine

15 [000311] Non-chiral analysis showing the substrate peak eluting at 2.6 min and 3.2 min. See Figures 8A-8E. The two peaks were present in all samples containing higher concentrations of the nitrile; the second peak is thought to be a product associated with the nitrile; it decreased with time and was no longer present once complete conversion to the product had taken place. The chromatogram shown in Figure 8A is a blank control,  
20 containing only nitrile and buffer; the samples were all diluted with water and solvent as explained in section 1 above. This was repeated for all samples discussed below. An enzymatic reaction sample is shown in the chromatogram in Figure 8B, with the product eluting at 0.4 min.

[000312] Of note in these chromatograms is the small solvent front peak eluting at 0.3 min.  
25 Further representation of this peak is given in the chromatogram shown in Figure 8C, in which a negative control consisting of cell lysate in buffer, was run. A very small peak coeluted with the product at 0.4 min. In the initial phase of the project, this peak was regarded as problematic, although the appropriate controls were run with each experiment for in order to maintain accuracy. In these experiments, the peak area resulting from the cell  
30 lysate, although it was relatively small, was subtracted from the peak areas of the product in the enzymatic reactions. Improvement of this analysis was obtained by further dilution of the samples and the use of lower injection volumes on the HPLC. Following the implementation



of these improvements, interference by this peak was shown to be minimal, as shown in the chromatogram illustrated in Fig. 6C.

[000313] The chiral analysis of phenylglycine is shown in chromatogram in Fig. 6D with the L-enantiomer eluting at 6 min and the D-enantiomer at 11 min. Good resolution between the two isomers was obtained. However, the column used was very sensitive and the characteristics of the column appeared to change over time, resulting in changes in the elution times of the acids. While this was easily detected by the use of the proper controls and standards, a greater problem existed in the coelution of the nitrile peak with the D-enantiomer (chromatogram shown in Fig. 6E). The cause of this coelution was unclear; however, it was easily detected by the use of appropriate standards; in addition, the UV spectrum of the acid was very distinctive, making the use of this tool effective in detecting the coelution. The problem was also easily resolved by adjusting the methanol content in the mobile phase.

(R)-2-chloromandelic acid

[000314] The HPLC analysis of chloromandelic acid and chloromandelonitrile offered many of the challenges associated with the analysis of the phenylglycine samples. From the chromatogram shown in Fig. 7A, which contains only chloromandelonitrile in buffer, it is evident that a peak eluted at the same time as the product in the chromatogram shown in Fig. 7B, which represents a chloromandelic acid standard. The contribution of the cell lysate to this peak was found to be small; it would appear that the greatest contribution to this peak was from the chloromandelonitrile, either from a breakdown product or a contaminant in the nitrile preparation. The peak area remained constant throughout each experiment and, using the appropriate controls, it was found that subtraction of the peak area from that of the product yielded sufficient accuracy. Many attempts were made to change the HPLC conditions so that the product peak eluted at a later time; however, these attempts were not successful. Chromatogram shown in Fig. 7C illustrates the appearance of product and the reduction of the substrate peaks.

[000315] The chiral analysis of chloromandelic acid was almost problem-free. The elution of a small peak at the same time as the (S)-enantiomer presented some concern (the peak at 2.4 min in chromatogram shown in Fig. 7D). However, once it was established that this peak was present in all the samples at the same level, including the blank control, and that it had a different UV spectrum to that of the chloromandelic acid peak, it was not regarded as a problem. Consequently, it was subtracted from the peak eluting at 2.4 min in each sample. The (R)-enantiomer eluted at 3 minutes.



(S)-phenyllactic acid

[000316] The analysis of phenyllactic acid was initially plagued with the same problems discussed for phenylglycine and 2-chloromandelic acid. However, in this case, adjustment of the solvent concentration in the nonchiral HPLC method led to a shift in the retention time of the acid, so that it no longer coeluted with the cell lysate peak. Following this, no problems were encountered with either the nonchiral or chiral methods. Representative nonchiral chromatograms of the product (1.9 min) and cyanohydrin substrate (3.7 min) are shown in Fig. 8A, while the chiral analysis of the acid is shown in Fig. 8B, with the L-enantiomer eluting at 2 min and the opposite enantiomer at 6 min.

L-2-methylphenylglycine

[000317] The analysis of methylphenylglycine was unproblematic, although the nonchiral method did not provide baseline separation between a cell lysate peak and the product peak, as shown in the chromatogram illustrated in Fig. 9A. The amino acid standard for this method was provided in the final stages of the project, thus minimizing the time for method development. In the chromatogram shown in Fig. 9A the amino acid elutes at 0.7 min and the aminonitrile at 5.0 min. Sufficient separation between the two initial peaks was obtained to allow the calculation of approximate conversion to product.

[000318] The chiral analysis of this compound provided good separation between the two enantiomers, as shown in the chromatogram illustrated in Fig. 9B. The L-enantiomer elutes at 5 min and the D-enantiomer at 8 min.

L-tert-leucine

[000319] For the nonchiral analysis of *t*-leucine, the cell lysate presented the most serious problem amongst the group of products for this project. This was compounded by the low spectroscopic properties of the amino acid, leading to difficulty in differentiating the product peak from the cell lysate. Good separation of the individual product enantiomers was obtained by chiral analysis as shown in Fig. 10A. During the primary screen, a small peak eluted at the same time as the L-amino acid standard in certain samples (see Fig. 10B) and was thought to be the amino acid. However, further development of the method and the use of the appropriate controls established that this peak was actually a cell lysate peak.

[000320] The aminonitrile eluted between the two *t*-leucine peaks, as shown in Fig. 10C; this chromatogram also shows the cell lysate peak at 4.8 min. The UV spectrum of the nitrile was distinct from that of the amino acid, making it easier to differentiate from the acid peaks.



L-hydroxynorleucine ((S)-2-amino-6-hydroxy hexanoic acid)

[000321] The chiral analysis of (S)-2-amino-6-hydroxy hexanoic acid was consistent and reliable. By contrast, the nonchiral method presented many problems, primarily as a result of non-separation between the nitrile and the acid peaks. Towards the latter half of the project, a method was developed and used successfully for the confirmation of activities. Prior to this, most of the analysis was performed using the chiral method; standard curves of the products were run in order to quantify the reactions. A representative chromatogram of (S)-2-amino-6-hydroxy hexanoic acid is shown in Figure 11A, with (S)-2-amino-6-hydroxy hexanoic acid eluting at 6 min. The aminonitrile was not detected by this method.

[000322] Separation of the individual 2-amino-6-hydroxy hexanoic acid enantiomers is shown in Fig. 11B. The L-enantiomer elutes first, at 2 min, followed by the D-enantiomer at 3 min. In Fig. 11C, an enzymatic sample is represented; the only area of slight concern is the negative peak preceding the elution of the L-enantiomer. However, it did not appear to interfere significantly with the elution of this enantiomer; method development did not eliminate the negative peak.

4-methyl-D-leucine and 4-methyl-L-leucine

[000323] For the detection of 4-methylleucine, the chiral HPLC method again proved more reliable. The combination of low activities, together with the low sensitivity of the method to the compound led to difficulties in detection using nonchiral HPLC. A 2.5 mM standard of the amino acid is shown in Fig. 12A, with a peak height of approximately 40 mAU; this was substantially lower than those detected for the aromatic compounds. Chromatogram in Fig. 12B shows an enzymatic sample, in which conversion was detected using the chiral HPLC method; while it is not clear, it would appear that the 4-methylleucine peak elutes at 2.7 min and is extremely low in both peak height and area. This peak did not appear in samples which were negative by chiral HPLC analysis.

[000324] The chiral analysis of 4-methyl-L-leucine and 4-methyl-D-leucine did not present any problems. The L-enantiomer eluted at 5 min and the D-enantiomer at 7 min, although some peak shift did occur, as a result of the sensitivity to the column, described in section (i) for phenylglycine. In chromatograms shown in Figs. 14C-14D, the separation of these amino acids is shown; the first sample represents an enzyme which produced both enantiomers and in the second sample, the enzyme preferentially hydrolyzed the L-enantiomer, with a small amount D-amino acid forming.



(S)-cyclohexylmandelic acid

[000325] Chromatograms of the standards for cyclohexylmandelic acid (Fig. 13A) and the corresponding nitrile (Fig. 13B) are shown. The acid eluted at 1.3 min, while the cyanohydrin was observed at 2.5 min. The peak eluting at 2.1 min is thought to be the cyclohexylphenylketone, as shown by the elution of a ketone standard at this point.

Example 12: An Enzyme Library Approach to Biocatalysis: Development of a Nitrilase Platform for Enantioselective Production of Carboxylic Acid Derivatives

[000326] Biocatalytic processes can offer unique advantages in transformations that are challenging to accomplish through conventional chemical methods (Wong, C.-H.;

Whitesides, G.M. *Enzymes in Synthetic Organic Chemistry*; Pergamon, New York, 1994; Drauz, K.; Waldmann, H., Roberts, S.M. Eds. *Enzyme Catalysis in Organic Synthesis*; VCH: Weinheim, Germany, 2nd ed., 2002). Nitrilases (EC 3.5.5.1) promote the mild hydrolytic conversion of organonitriles directly to the corresponding carboxylic acids (Kobayashi, M.; Shimizu, S. *FEMS Microbiol. Lett.* 1994, 120, 217; Bunch, A.W. In *Biotechnology*; Rehm, H.-J.; Reed, G.; Puhler, A.; Stadler, P., Eds.; Wiley-VCH: Weinheim, Germany, Vol. 8a, Chapter 6, pp 277-324; Wieser, M.; Nagasawa, T. In *Stereoselective Biocatalysis*; Patel, R.N., Ed.; Marcel Dekker: New York, 2000, Chapter 17, pp 461-486.) Fewer than fifteen microbially-derived nitrilases have been characterized and reported to date. (Harper, D.B. *Int. J. Biochem.* 1985, 17, 677; Levy-Schil, S.; Soubrier, F.; Crutz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D. *Gene* 1995, 161, 15; Yu, F. 1999, US Patent 5872000; Ress-Loschke, M.; Friedrich, T.; Hauer, B.; Mattes, R.; Engels, D. PCT Appl. WO 00/23577, April 2000.). Several nitrilases previously have been explored for the preparation of single-enantiomer carboxylic acids, although little progress has been made in the development of nitrilases as viable synthetic tools. This application describes the discovery of a large and diverse set of nitrilases and herein demonstrate the utility of this nitrilase library for identifying enzymes that catalyze efficient enantioselective production of valuable hydroxy carboxylic acid derivatives.

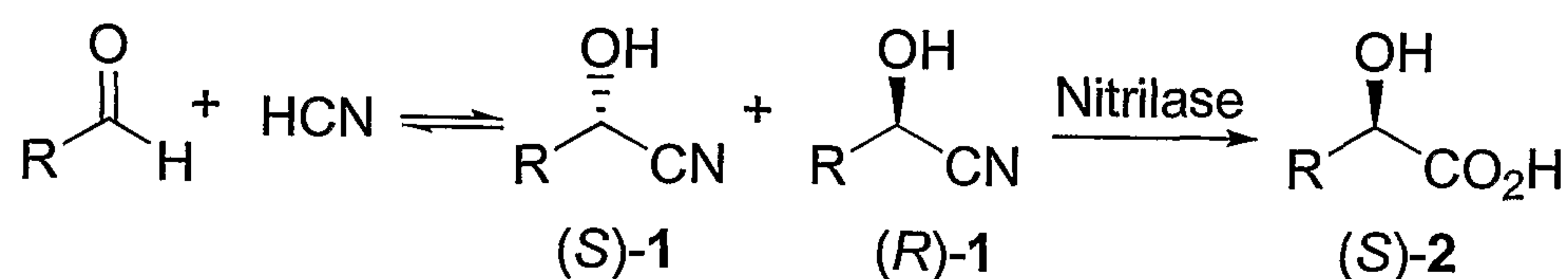
[000327] In an effort to access the most diversified range of enzymes that can be found in Nature, we create large genomic libraries by extracting DNA directly from environmental samples that have been collected from varying global habitats. (For a description of these methods, see: Short, J.M. *Nature Biotech.* 1997, 15, 1322; Handelsman, J.; Rondon, M.J.; Brady, S.F.; Clardy, J.; Goodman, R.M. *Chem. Biol.* 1998, 5, R245; Henne, A.; Daniel, R.; Schmitz, R.A.; Gottschalk, G. *Appl. Environ. Microbiol.* 1999, 65, 3901.). We have



established a variety of methods for identifying novel activities through screening mixed populations of uncultured DNA. (Robertson, D.E.; Mathur, E.J.; Swanson, R.V.; Marrs, B.L.; Short, J.M. *SIM News* **1996**, 46, 3; Short, J.M. US Patent 5,958,672, **1999**; Short J.M. US Patent 6,030,779, **2000**.) Through this approach, nearly 200 new nitrilases have been discovered and characterized. (For a concise description of the studies, see Materials and Methods section below.) All nitrilases were defined as unique at the sequence level and were shown to possess the conserved catalytic triad Glu-Lys-Cys which is characteristic for this enzyme class. (Pace, H.; Brenner, C. *Genome Biology* **2001**, 2, 0001.1-0001.9.) Each nitrilase in our library was overexpressed and stored as a lyophilized cell lysate in order to facilitate rapid evaluation of the library for particular biocatalytic functions.

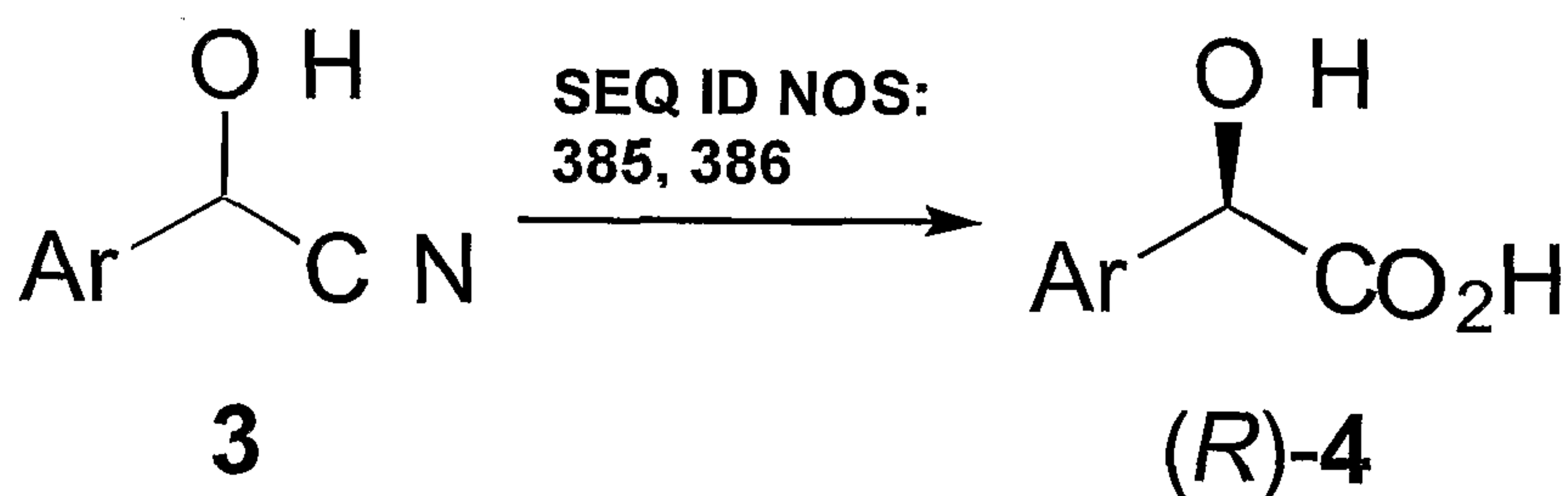
**[000328]** The initial investigations focused upon the efficacy of nitrilases for production of  $\alpha$ -hydroxy acids **2** formed through hydrolysis of cyanohydrins **1**. Cyanohydrins are well-documented to racemize readily under basic conditions through reversible loss of HCN. (Inagaki, M.; Hiratake, J.; Nishioka, T.; Oda, J.; *J. Org. Chem* **1992**, 57, 5643. (b) van Eikeren, P. US Patent 5,241,087, 1993.) Thus, a dynamic kinetic resolution process is possible whereby an enzyme selectively hydrolyzes only one enantiomer of **1**, affording **2** in 100% theoretical yield and with high levels of enantiomeric purity.

**[000329]** One important application of this type involves commercial production of (*R*)-mandelic acid from mandelonitrile. (Ress-Loschke, M.; Friedrich, T.; Hauer, B.; Mattes, R.; Engels, D. PCT Appl. WO 00/23577, April 2000; Yamamoto, K.; Oishi, K.; Fujimatsu, I.; Komatsu, K. *Appl. Environ. Microbiol.* **1991**, 57, 3028; Endo, T.; Tamura, K. US Patent 5,296,373, March 1994.) Mandelic acid and derivatives find broad use as intermediates and resolving agents for production of many pharmaceutical and agricultural products. (Coppola, G.M.; Schuster, H.F. *Chiral  $\alpha$ -Hydroxy Acids in Enantioselective Synthesis*; Wiley-VCH: Weinheim, Germany: 1997.) However, the few known nitrilases derived from cultured organisms have not been found useful for efficient and selective hydrolysis of analogous substrates.



**[000330]** The nitrilase library was screened for activity and enantioselectivity in the hydrolysis of mandelonitrile (3a, Ar = phenyl) to mandelic acid. Preliminary results revealed



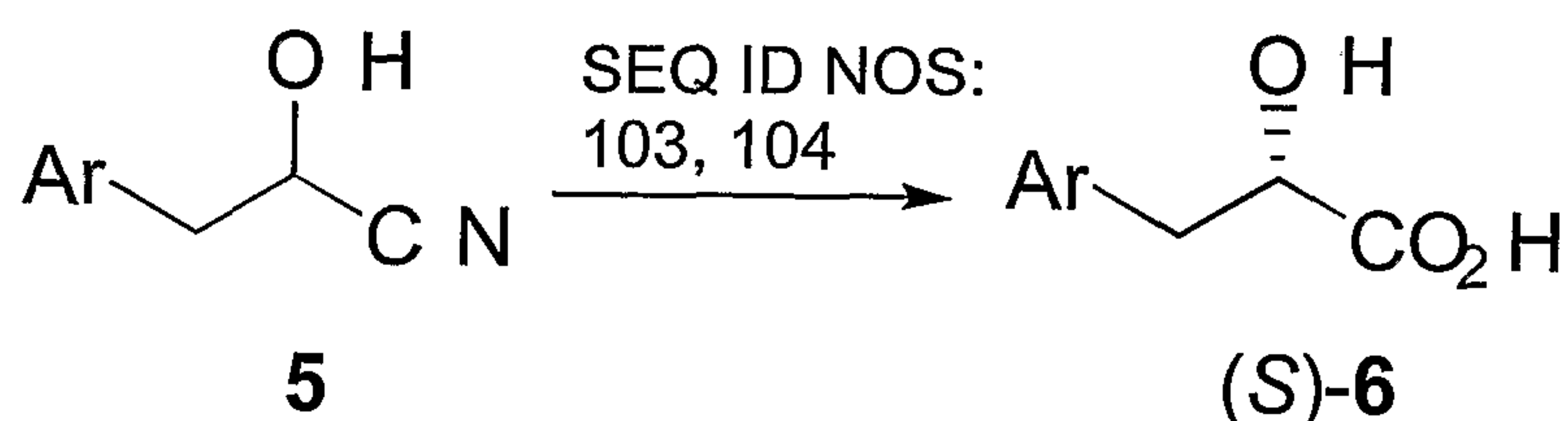


that 27 enzymes afforded mandelic acid in >90% ee. One enzyme, SEQ ID NOS:385, 386, was studied in greater detail and was found to be very active for hydrolysis of mandelonitrile. Under standard conditions using 25 mM 3a and 0.12 mg/mL enzyme in 10% MeOH (v/v) 0.1 M phosphate buffer at 37°C and pH 8, (R)-mandelic acid was formed quantitatively within 10 min and with 98% ee. To confirm synthetic utility, the reaction was performed using 1.0 g 3a (50 mM) and 9 mg nitrilase (0.06 mg/mL nitrilase I); after 3 h (R)-mandelic acid was isolated in high yield (0.93 g, 86%) and again with 98% ee.

(a) Reactions were conducted under standard conditions (see text). Reaction time for complete conversion to 4 was 1-3 h. Entries 8-9 were conducted at pH 9 and 5 mM substrate concentration. (b) Specific activities were measured at 5 min transformation timepoints and are expressed as  $\mu\text{mol mg}^{-1} \text{ min}^{-1}$ . (c) TOF = turnover frequency, mol product/mol catalyst/sec. (d) Enantioselectivities were determined by chiral HPLC analysis. Hydroxy acids were isolated and absolute configurations were determined to be (R) in all cases.

**[000331]** The substrate scope of SEQ ID NOS:385, 386 was next explored. As shown in Table 13, a broad range of mandelic acid derivatives as well as aromatic and heteroaromatic analogues (4) may be prepared through this method. SEQ ID NOS:385, 386 tolerates aromatic ring substituents in the *ortho*-, *meta*-, and *para*-positions of mandelonitrile derivatives and products of type 4 were produced with high enantioselectivities. Other larger aromatic groups such as 1-naphthyl and 2-naphthyl also are accommodated within the active site, again affording the acids 4 with high selectivity (Table 13, entries 8-9). Finally, 3-pyridyl and 3-thienyl analogues of mandelic acid were prepared readily using this process (Table 13, entries 10-11). This is the first reported demonstration of a nitrilase that affords a range of mandelic acid derivatives and heteroaromatic analogues of type 4. High activity on the more sterically encumbered *ortho*-substituted and 1-naphthyl derivatives is particularly noteworthy.





[000332] We next examined the preparation of aryllactic acid derivatives **6** through hydrolysis of the corresponding cyanohydrins **5**. Phenyllactic acid and derivatives serve as versatile building blocks for the preparation of numerous biologically active compounds.

(Coppola, G.M.; Schuster, H.F. *Chiral  $\alpha$ -Hydroxy Acids in Enantioselective Synthesis*;

5 Wiley-VCH: Weinheim, Germany: 1997.) Upon screening our nitrilase library against the parent cyanohydrin **5a** (Ar = phenyl), we found several enzymes that provided **6a** with high enantiomeric excess. One enzyme, SEQ ID NOS: 103, 104, was further characterized. After optimization, SEQ ID NOS:103, 104, was shown to provide (*S*)-phenyllactic acid (**6a**) with complete conversion (50 mM) and very high enantioselectivity (98% ee) over 6 h. The  
 10 highest enantioselectivity previously reported for biocatalytic conversion of **5** to **6** was 75% ee achieved through a whole cell transformation using a *Pseudomonas* strain. (Hashimoto, Y.; Kobayashi, E.; Endo, T.; Nishiyama, M.; Horinouchi, S. *Biosci. Biotech. Biochem.* 1996, 60, 1279.)

Table 7. Nitrilase II-catalyzed production of aryllactic acid derivatives and analogues **6**<sup>a</sup>

| Entry | Ar in <b>6</b>                     | Spec. Act. <sup>b</sup> | TOF <sup>c</sup> | % ee <sup>d</sup> |
|-------|------------------------------------|-------------------------|------------------|-------------------|
| 1     | C <sub>6</sub> H <sub>5</sub>      | 25                      | 16               | 99                |
| 2     | 2-Me-C <sub>6</sub> H <sub>5</sub> | 160                     | 100              | 95                |
| 3     | 2-Br-C <sub>6</sub> H <sub>5</sub> | 121                     | 76               | 95                |
| 4     | 2-F-C <sub>6</sub> H <sub>5</sub>  | 155                     | 97               | 91                |
| 5     | 3-Me-C <sub>6</sub> H <sub>5</sub> | 21                      | 13               | 95                |
| 6     | 3-F-C <sub>6</sub> H <sub>5</sub>  | 22                      | 14               | 99                |



WO 03/097810

PCT/US03/15712

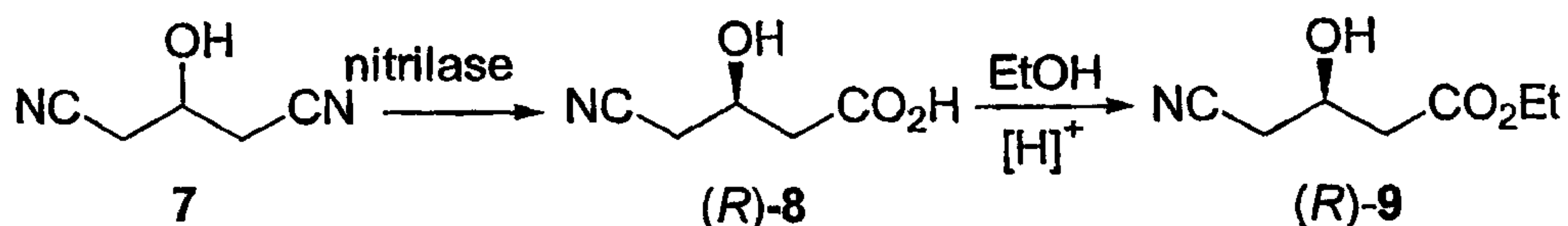
|    |            |      |     |    |
|----|------------|------|-----|----|
| 7  | 1-naphthyl | 64   | 40  | 96 |
| 8  | 2-pyridyl  | 10.5 | 6.6 | 99 |
| 9  | 3-pyridyl  | 11.6 | 7.2 | 97 |
| 10 | 2-thienyl  | 3.4  | 2.1 | 96 |
| 11 | 3-thienyl  | 2.3  | 1.4 | 97 |

(a) Reaction conditions as in Table 13, except 0.016 mg/mL nitrilase was used. Full conversion to **6** was observed within 6 h. (b)-(d) See Table 13. The absolute configuration was determined to be (*S*) for phenyllactic acid and entries 2-11 were assigned (*S*) based upon identical chiral HPLC peak elution order.

5 [000333] *Ortho* and *meta* substituents appear to be tolerated well by nitrilase II, with *ortho* substituted derivatives surprisingly being converted with higher rates relative to the parent substrate **5a**. Novel heteroaromatic derivatives, such as 2-pyridyl-, 3-pyridyl, 2-thienyl- and 3-thienyllactic acids, were prepared with high conversions and enantioselectivities (entries 8-11). Unexpectedly, *para* substituents greatly lowered the rates  
10 of these reactions, with full conversion taking over two weeks under these conditions.

[000334] The final transformation that we examined was desymmetrization of the readily available prochiral substrate 3-hydroxyglutarylnitrile (**7**) (Johnson, F.; Panella, J.P.; Carlson, A.A. *J. Org. Chem.* **1962**, 27, 2241) to afford hydroxy acid (*R*)-**8** which, once esterified to (*R*)-**9**, is an intermediate used in the manufacture of the cholesterol-lowering  
15 drug LIPITOR<sup>TM</sup>. Previously reported attempts to use enzymes for this process were unsuccessful and **8** was produced with low selectivity (highest: 22% ee) and the undesired (*S*)-configuration. (Crosby, J.A.; Parratt, J.S.; Turner, N.J. *Tetrahedron: Asymmetry* **1992**, 3, 1547; Beard, T.; Cohen, M.A.; Parratt, J.S.; Turner, N.J. *Tetrahedron: Asymmetry* **1993**, 4, 1085; Kakeya, H.; Sakai, N.; Sano, A.; Yokoyama, M.; Sugai, T.; Ohta, H. *Chem. Lett.* **1991**,  
20 1823.)





[000335] The nitrilase library was screened and unique enzymes were discovered and isolated that provided the required product (R)-8 with high conversion (>95%) and >90% *ee*. Using one of the (R)-specific nitrilases, this process was operated on a 1.0 g scale (240 mM 7, 30 mg enzyme, 22°C, pH 7) and after 22 h, (R)-8 was isolated in 98% yield and 95% *ee*. Interestingly, the same screening program also identified nitrilases that afford the opposite enantiomer (S)-8 with 90-98% *ee*. Thus, the extensive screen of biodiversity has uncovered enzymes that provide ready access to either enantiomer of the intermediate 8 with high enantioselectivities. Our discovery of the first enzymes that furnish (R)-8 underscores the advantage of having access to a large and diverse library of nitrilases.

[000336] By plumbing our environmental genomic libraries created from uncultured DNA, we have discovered a large array of novel nitrilases. This study has revealed specific nitrilases that furnish mandelic and aryllactic acid derivatives, as well as either enantiomer of 4-cyano-3-hydroxybutyric acid in high yield and enantiomeric excess.

#### 15 Procedures and Analytical Data:

[000337] Hydroxyglutarylnitrile was purchased from TCI America and used as received. Amino acids used for the preparation of aryl lactic acid standards were purchased from PepTech (Cambridge, MA). (R)-3-hydroxy-4-cyanobutyric acid was obtained from Gateway Chemical Technology (St. Louis, MO). Both (R)- and (S)- mandelic acid and (R)- and (S)- phenyl lactic acid standards were purchased from Sigma Aldrich. All other reagents were purchased from Sigma Aldrich and utilized without further purification. Silica Gel, 70-230 mesh, 60 Å, purchased from Aldrich, was used for chromatographic purifications. All <sup>1</sup>H NMRs and <sup>13</sup>C NMRs were run on Bruker model AM-500 machines, set at room temperature, 500 MHz and 125MHz respectively for <sup>1</sup>H and <sup>13</sup>C. Mass analyses and unit mass resolution was achieved by flow injection analysis (FIA) using a Perkin-Elmer Sciex\* API-4000 TURBOION™ Spray LC/MS/MS system. The LC flow was provided by Shimadzu LC-10Advp pumps, with 0.05% acetic acid and MeOH. Injections were accomplished via a Valco\* injector valve. The HPLC analysis was done on an Agilent\* 1100 HPLC with Astec's Chirobiotic\*R column (100 x 4.6 mm, cat no. 13022 or 150 x 4.6 mm, cat no. 13023) or Daicel's Chiralcel\* OD column (50 x 4.6 mm, cat no. 14022) and the DAD detector set at 210,

\*Trade-mark



220, 230, and 250 nm. For specific rotations, a Perkin Elmer Model 341 Polarimeter was used, set at 589 nm, Na lamp, at room temperature, with a 100 mm path length cell.

Concentrations for specific rotation are reported in grams per 100 mL of solvent.

Microbiology techniques were executed in accordance to published protocols. (Sambrook, J.

- 5 Fritsch, EF, Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed.), Cold Spring Harbor Laboratory Press, Plainview NY.) Glycolic acid products were isolated and absolute configurations were determined to be (*R*) in all cases by comparison with literature optical rotation data on configurationally defined compounds except for (-)-3-pyridylglycolic acid, which to our knowledge is not known as a single enantiomer. (For mandelic, 2-  
10 chloromandelic, 2-methyl mandelic, 3-chloromandelic, 3-bromomandelic and 4-fluoromandelic acid see Hoover, J.R. E.; Dunn, G. L.; Jakas, D.R.; Lam, L.L.; Taggart, J. J.; Guarini, J.R.; Phillips, L. *J. Med. Chem.* **1974**, 17(1), 34-41; For 2-bromo mandelic acid see Collet, A.; Jacques, J.; *Bull. Soc. Chem. Fr.* **1973**, 12, 3330-3331; For 1- and 2-naphthylglycolic acid see Takahashi, I; Y. Aoyagi, I. Nakamura, Kitagawa, A., Matsumoto,  
15 K., Kitajima, H. Isa, K. Odashima, K. Koga, K. *Heterocycles* **1999**, 51(6), 1371-88; For 3-thienylglycolic acid Gronowitz, S. *Ark. Kemi*, **1957**, 11, 519-525.)

- [000338] For the aryl lactic acid products, absolute configuration was established to be (*S*) for phenyl lactic acid by comparison with literature optical rotation and for all other phenyl lactic acid products, absolute configurations were predicted based upon elution order  
20 using chiral HPLC. Absolute configuration for 3-hydroxy-4-cyano-butanoic acid was established by derivatization to (*R*)-(-)-Methyl (3-O-[benzoyl]-4-cyano)-butanoate and comparison to literature optical rotation data on configurationally defined compound. (3. Beard, T. Cohen, M. A. Parratt, J.S. Turner, N. J. *Tetrahedron:Asymm.* 4(6), **1993**, 1085-1104.)

## 25 Nitrilase Discovery and Characterization Methods:

### 1. Nitrilase Selection.

- [000339] An *Escherichia coli* screening host strain, SEL700, was optimized for nitrilase selections on a nitrile substrate. An  $Abs_{600nm} = 1$ , resuspension of SEL700 screening host in 10 mM  $MgSO_4$  was infected with kanamycin-resistant environmental DNA library for 45  
30 minutes at 37°C, such that complete screening coverage of the library was achieved. Infected cells, now denoted by kanamycin resistance, were plated on kanamycin LB plates and allowed to grow overnight at 30°C. Titer plates were also made to determine infection



efficiency. Cells were pooled, washed, and resuspended the next morning with 10 mM MgSO<sub>4</sub>. Transformed clones were inoculated into M9 media (without nitrogen) with 10 mM of nitrile substrate. M9 media consisted of 1X M9 salts (NH<sub>4</sub>Cl omitted), 0.1mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 0.2 % glucose, and approximately 10 mM of a nitrile selection substrate. The selection cultures were then incubated at 30°C, shaking at 200 rpm, for up to five weeks.

Positive nitrilase cultures were identified by growth, due to positive clone's ability to hydrolyze nitrile substrate. Positive clones were isolated by streaking out a selection culture with growth and subsequent secondary culturing of isolated colonies in the same defined media. The DNA from any positive secondary cultures exhibiting re-growth was then isolated and sequenced to confirm discovery of a nitrilase gene and to establish the unique nature of that gene.

## 2. Nitrilase Biopanning.

[000340] Traditional filter lift hybridization screening protocols are limited to libraries with approximately 10<sup>6</sup> to 10<sup>7</sup> members. Attempting to screen one library would require approximately 5,000 filter lifts. Therefore, solution phase and other biopanning formats have been developed for ultra high throughput sequence based screening permitting rapid screening of up to 10<sup>8</sup> member environmental libraries. In the solution format, the DNA from a large number of library clones is mixed with tagged molecules of interest under conditions which promote hybridization. The tagged clones and hybridized DNA are then removed from solution and washed at some level of stringency to remove clones which do not have sequence identity with the probe. The hybridized DNA is then eluted and recovered. Clones of interest are sequenced and cloned to provide enzyme activities of interest. This method has been demonstrated to achieve up to 1,000-fold enrichment per round for sequences of interest.

## 3. High Throughput Nitrilase Activity Assay.

[000341] Activity assays were conducted using 25 mM (~3 mg/mL) substrate, 0.1 mg/mL nitrilase in 0.25 mL of assay solution. Assay solutions consisted of 0-10% (v/v) MeOH in 0.1 M sodium phosphate buffer solution at pH 7 to 9 and temperatures 37°C or 22°C. Specific activities were measured at 5 min transformation time point, unless otherwise noted, and are expressed in units  $\mu\text{mol mg}^{-1} \text{ min}^{-1}$ . Enantiomeric excess and conversion rates were determined by high throughput HPLC analysis comparing enzyme product concentration to standard curves of racemic acid products. Analytical conditions for the products are tabulated below.



## Analytical Methods:

|      | Acid Product                            | Column                        | Liquid Chromatography Method                                              | Retention Times of enantiomers (min) |
|------|-----------------------------------------|-------------------------------|---------------------------------------------------------------------------|--------------------------------------|
| 1.1  | mandelic acid                           | Chirabiotic R<br>100 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.4 ( <i>S</i> ); 2.9 ( <i>R</i> )   |
| 1.2  | 2-Cl-mandelic acid                      | Chirabiotic R<br>100 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.3 ( <i>S</i> ); 2.9 ( <i>R</i> )   |
| 1.3  | 2-Br-mandelic acid                      | Chirabiotic R<br>100 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.8; 4.0                             |
| 1.4  | 2-CH <sub>3</sub> -mandelic acid        | Chirabiotic R<br>100 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 3.1; 3.8                             |
| 1.5  | 3-Cl-mandelic                           | Chirabiotic R<br>100 x 4.6 mm | 10%[0.5% AcOH], 90% CH <sub>3</sub> CN<br>1 ml/min                        | 3.1; 3.8                             |
| 1.6  | 3-Br-mandelic                           | Chirabiotic R<br>100 x 4.6 mm | 10%[0.5% AcOH], 90% CH <sub>3</sub> CN<br>1 ml/min                        | 3.3; 3.9                             |
| 1.7  | 4-F-mandelic                            | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 3.7; 4.8                             |
| 1.8  | 1-naphthylglycolic acid                 | Chirabiotic R<br>100 x 4.6 mm | 4%[0.5% AcOH], 96% CH <sub>3</sub> CN<br>1 ml/min                         | 3.1; 3.7                             |
| 1.9  | 2-naphthylglycolic acid                 | Chirabiotic R<br>100 x 4.6 mm | 4%[0.5% AcOH], 96% CH <sub>3</sub> CN<br>1 ml/min                         | 3.7; 4.7                             |
| 1.10 | 3-pyridylglycolic acid                  | Chirabiotic R<br>100 x 4.6 mm | 5% [0.5% AcOH], 65% H <sub>2</sub> O,<br>30% CH <sub>3</sub> CN, 2 ml/min | 4.4; 5.5                             |
| 1.11 | 3-thienylglycolic acid                  | Chirabiotic R<br>100 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>2 ml/min                        | 1.4; 2.5                             |
| 2.1  | phenyl lactic acid                      | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.8 ( <i>S</i> ); 4.0 ( <i>R</i> )   |
| 2.2  | 2-methylphenyl lactic acid              | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.5; 2.8                             |
| 2.3  | 2-bromophenyl lactic acid               | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.8; 3.2                             |
| 2.4  | 2-fluorophenyl lactic acid              | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.6; 2.9                             |
| 2.5  | 3-methylphenyl lactic acid              | Chirabiotic R<br>150 x 4.6 mm | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.4; 3.2                             |
| 2.6  | 3-fluorophenyl lactic acid              | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.8; 3.6                             |
| 2.7  | 1-naphthyllactic acid                   | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.7; 3.1                             |
| 2.8  | 2-pyridyllactic acid                    | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.5; 2.9                             |
| 2.9  | 3-pyridyllactic acid                    | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 2.9; 3.6                             |
| 2.10 | 2-thienyllactic acid                    | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 3.6; 4.6                             |
| 2.11 | 3-thienyllactic acid                    | Chirabiotic R<br>150 x 4.6mm  | 20%[0.5% AcOH], 80% CH <sub>3</sub> CN<br>1 ml/min                        | 3.5; 4.6                             |
|      | Methyl(3-O [benzoyl]-4-cyano)-butanoate | Daicel OD<br>50 x 4.6 mm      | 5% isopropanol, 95% hexane<br>1 ml/min                                    | 4.5 ( <i>R</i> ); 5.4( <i>S</i> )    |

Cyanohydrin (Substrate) Synthesis:

- 5 [000342] Mandelonitrile Synthesis Method A: Acetone cyanohydrin (685  $\mu$ L, 7.5 mmol), aldehyde (5 mmol), and catalytic DIEA (13  $\mu$ L, 0.075 mmol) were mixed at 0 °C. The



reactions were stirred on ice for 45 minutes. To drive the equilibrium toward the product, acetone was removed *in vacuo*. Subsequently, crude reactions were acidified with H<sub>2</sub>SO<sub>4</sub> (3  $\mu$ L) and stored at -20°C. TLC was used to monitor reaction progress (3:1 hexane/ethylacetate (EtOAc)).

- 5 [000343] Mandelonitrile Synthesis Method B: To a solution of KCN (358 mg, 5.5 mmol) in MeOH (1 mL) at 0°C was added aldehyde (5 mmol) and acetic acid (315  $\mu$ L, 5.5 mmol). After stirring for one hour on ice, MeOH was removed *in vacuo*, and the crude mixture was partitioned using EtOAc and H<sub>2</sub>O. The organic fraction was retained and concentrated *in vacuo*. TLC analysis was used to monitor reaction progress (3:1 Hexanes/EtOAc).
- 10 [000344] Aryl Acetaldehyde Cyanohydrin: Arylacetic acid (50 mmol) was dissolved in 50 ml anhydrous tetrahydrofuran (THF) in a two-neck 500 ml round-bottom flask under N<sub>2</sub>(g) atmosphere. To this solution cooled to 0 °C, under vigorous mixing, was added slowly 105 mmol of hexylchloroborane-dimethyl sulfide (2.55 M in methylene chloride). The reaction was allowed to proceed overnight. Excess acetic acid (10 ml) was added to quench and
- 15 acidify the reaction followed by the addition 10 ml water. After stirring at room temperature for 1 hour, solvent was removed *in vacuo* and the residue was dissolved in 100 ml water and extracted with 200 ml EtOAc. The EtOAc layer was dried over sodium sulfate, filtered and then concentrated *in vacuo*. Subsequently, 60 mmol of KCN, followed by 100 ml methanol was added to the residue. The solution was then cooled to 0 °C and acetic acid (60 mmol)
- 20 added. The reaction was stirred for 1-2 hours after all KCN dissolved. Solvents were removed *in vacuo* and residue was dissolved in 100 ml water and 200 ml EtOAc. The aqueous layer was extracted with EtOAc one more time. Combined EtOAc extracts were washed with saturated brine and dried over sodium sulfate, filtered and then concentrated *in vacuo* to obtain crude cyanohydrin product. The cyanohydrin was purified by silica-gel column
- 25 (hexane/EtOAc), as necessary.

[000345] 2-chloro mandelonitrile: <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz)  $\delta$  7.69 (m, 1H), 7.41 (m, 1H), 7.36 (m, 2H), 5.84 (s, 1H), 3.07 (br, 1H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz)  $\delta$  132.89, 132.73, 131.22, 130.19, 128.48, 127.84, 118.24, 60.87. MS calc'd for [C<sub>8</sub>H<sub>6</sub>ClNO] 167.01 found 167.9 (LC-MS +).

- 30 [000346] 2-bromomandelonitrile: <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz)  $\delta$  7.72 (d, 1H, *J*= 6.58), 7.62 (d, 1H, *J*= 8.35), 7.43 (t, 1H, *J*= 8.42), 7.30 (t, 1H, *J*= 7.00), 5.85 (s, 1H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz)  $\delta$  134.550, 133.584, 131.564, 128.819, 128.535, 122.565, 118.153, 63.379.



[000347] 2-methylmandelonitrile:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$ : 7.60 (d, 1H,  $J = 7.4$ ), 7.23-7.35 (m, 3H), 5.66 (s, 1H), 2.44 (s, 3H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 298 K, 125 MHz)  $\delta$ : 136.425, 133.415, 131.450, 130.147, 127.204, 126.894, 118.952, 18.916. MS calc'd for  $[\text{C}_9\text{H}_9\text{NO}]$  147.07, found 147.2 (ESI +).

5 [000348] 3-chloromandelonitrile:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.55 (s, 1H), 7.43-7.37 (m, 3H), 5.54 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  137.183, 135.480, 130.718, 130.303, 127.047, 124.891, 118.395, 63.156. MS calc'd for  $[\text{C}_8\text{H}_6\text{ClNO}]$  167.01 found 167.9 (LC-MS +).

10 [000349] 3-bromomandelonitrile:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.69 (s, 1H), 7.56 (d,  $J = 6.2$  Hz, 1H), 7.45 (d,  $J = 5.5$  Hz, 1H), 7.32 (t,  $J = 6.4$  Hz, 1H), 5.53 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  137.376, 133.201, 130.934, 129.208, 125.359, 123.380, 118.458, 63.006. MS calc'd for  $[\text{C}_8\text{H}_6\text{BrNO}]$  212.0 found 211.9 (LC-MS +).

15 [000350] 4-fluoromandelonitrile:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  5.54 (s, 1H), 7.13 (m, 2H), 7.51-7.53 (m, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  63.02, 116.44, 118.97, 128.90, 131.54, 132.51, 162.575.

[000351] 4-chloromandelonitrile:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.47 (d,  $J = 7.0$  Hz, 2H), 7.42 (d,  $J = 7.0$  Hz, 2H), 5.53 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  136.209, 133.845, 129.647, 128.232, 118.630, 63.154. MS calc'd for  $[\text{C}_8\text{H}_6\text{ClNO}]$  167.01 found 167.9 (LC-MS +)

20 [000352] 1-naphthyl cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  8.14 (d, 1H,  $J = 8.5$ ), 7.92 (t, 2H,  $J = 6.1$ ), 7.82 (d, 1H,  $J = 5.7$ ), 7.62 (t, 1H,  $J = 6.1$ ), 7.56 (t, 1H,  $J = 6.1$ ), 7.50 (t, 1H,  $J = 6.1$ ), 6.18 (s, 1H);  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  137.0, 135.7, 134.2, 131.1, 129.2, 127.5, 126.7, 125.8, 125.3, 123.1, 119.0, 62.4; MS calc'd for  $[\text{C}_{12}\text{H}_9\text{O}]$  183.21, found 183.2 (ESI +).

25 [000353] 2-naphthyl cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  8.03 (s, 1H), 7.92 (d, 1H,  $J = 8.6$ ), 7.87-7.91 (m, 2H), 7.61 (dd, 1H,  $J = 6.7, 1.2$ ), 7.55-7.60 (m, 2H), 5.72 (s, 1H);  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  134.9, 133.9, 132.7, 129.6, 128.6, 128.0, 127.4, 127.2, 126.4, 123.9, 118.9, 64.1; MS calc'd for  $[\text{C}_{12}\text{H}_9\text{O}]$  183.21, found 183.2 (ESI +).

30 [000354] 3-pyridyl cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$ : 8.62 (d, 1H,  $J = 1.8$ ), 8.57 (d, 1H,  $J = 5.1$ ), 7.94 (d, 1H,  $J = 8.1$ ), 7.41 (dd, 1H,  $J = 8.1, 5.1$ ), 5.64 (s, 1H);  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  149.921, 147.355, 135.412, 133.044, 124.443, 118.980, 61.085. MS calc'd for  $[\text{C}_7\text{H}_6\text{N}_2\text{O}]$  134.05, found 135.2 (ESI +).



[000355] 3-thienyl cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.45 (d,  $J = 2.2$  Hz 1H), 7.56 (dd,  $J = 6.2$  Hz, 1H), 7.45 (d,  $J = 5.5$  Hz, 1H), 7.32 (t,  $J = 6.4$  Hz, 1H), 5.53 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  137.376, 133.201, 130.934, 129.208, 125.359, 123.380, 118.458, 63.006. MS calc'd for  $[\text{C}_6\text{H}_5\text{NOS}]$  139.01 found 139.9 (LC-MS +).

5 [000356] phenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.34 (m, 5H), 4.64 (t,  $J = 6.75$  Hz, 1H), 3.11 (d,  $J = 6.75$  Hz, 2H), 2.75 (br, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  133.96, 129.91, 129.16, 128.08, 119.47, 62.33, 41.55.

[000357] 2-methylphenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.11 (m, 4H), 4.61 (t,  $J = 6.62$  Hz, 1H), 3.12 (d,  $J = 6.62$  Hz, 2H), 2.14 (s, 3H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  136.94, 136.47, 132.57, 130.48, 127.61, 125.75, 120.11, 62.95, 44.73 MS calc'd for  $[\text{C}_{10}\text{H}_{11}\text{NO}]$ : 161.08, found 162.2 (M+Na, ESI +)

10

[000358] 2-bromophenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.20 (m, 4H), 4.78 (t,  $J = 6.5$  Hz, 1H), 3.26 (d,  $J = 6.5$  Hz, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz)  $\delta$  133.93, 132.82, 131.72, 129.21, 128.12, 124.86, 119.41, 63.02, 44.89.

15 [000359] 2-fluorophenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.2 (m, 2H), 7.02 (m, 2H), 4.50 (dd,  $J = 4.62$  Hz,  $J = 7.88$  Hz, 1H), 3.23 (dd,  $J = 4.62$  Hz,  $J = 14.12$  Hz, 1H), 2.97 (dd,  $J = 7.88$  Hz,  $J = 14.12$  Hz, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  132.18, 131.52, 129.66, 129.03, 128.07, 124.05, 115.8, 63.02, 44.79 MS calc'd for  $[\text{C}_9\text{H}_8\text{FNO}]$  165.06, found 164.2 (ESI +).

20 [000360] 3-methylphenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.18 (m, 1H), 7.02 (m, 3H), 4.54 (dd,  $J = 4.62$  Hz,  $J = 8$  Hz, 1H), 3.06 (dd,  $J = 4.62$  Hz,  $J = 14.38$  Hz, 1H), 2.83 (dd,  $J = 8$  Hz,  $J = 14.38$  Hz, 1H), 2.36 (s, 3H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  176.25, 138.18, 136.0, 130.97, 128.93, 127.68, 126.58, 76.42, 34.29, 37.69 MS calc'd for  $[\text{C}_{10}\text{H}_{12}\text{O}_3]$  180.08, found 180.0 (ESI +).

25 [000361] 3-fluorophenyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.18 (m, 2H), 6.95 (m, 2H), 4.44 (dd, 1H), 3.11 (dd, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  130.40, 125.53, 124.85, 116.92, 114.87, 114.50, 119.77, 61.97, 41.27.

[000362] 1-naphthyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  8.07 (m, 1H), 7.86 (m, 1H), 7.74 (m, 1H), 7.41 (m, 4H), 4.20 (t,  $J = 7$  Hz, 1H), 3.33 (d,  $J = 6.8$  Hz, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  177.7, 140.31, 129.74, 129.24, 128.92, 128.26, 127.84, 125.63, 124.53, 124.05, 123.42, 70.58, 38.0 MS calc'd for  $[\text{C}_{13}\text{H}_{11}\text{NO}]$  197.08, found 197.1 (ESI +).

30



[000363] 2-pyridyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  8.50 (m, 1H), 7.85 (m, 1H), 7.48 (m, 1H), 7.34 (m, 1H), 4.42 (m, 1H), 3.19 (dd,  $J = 3.5$  Hz,  $J = 13.7$  Hz, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  157.44, 145.69, 140.24, 126.96, 126.16, 122.99, 60.30, 42.60 MS calc'd for  $[\text{C}_8\text{H}_8\text{N}_2\text{O}]$  148.06, found 149.1 (ESI +).

5 [000364] 3-pyridyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  8.62 (d, 1H,  $J = 1.8$ ), 8.57 (d, 1H,  $J = 5.1$ ), 7.94 (d, 1H,  $J = 8.1$ ), 7.41 (dd, 1H,  $J = 8.1$ , 5.1), 5.64 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$ : 149.921, 147.355, 135.412, 133.044, 124.443, 118.980, 61.085. Exact Mass calculated for  $[\text{C}_7\text{H}_6\text{N}_2\text{O}]$ : 134.05, found: 135.2 (ESI +).

10 [000365] 2-thienyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.1 (m, 1H), 6.9 (m, 1H), 6.8 (m, 1H), 4.11 (t,  $J = 7.0$  Hz, 1H), 2.86 (d,  $J = 7.0$  Hz, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  127.68, 127.41, 125.58, 124.60, 118.70, 63.25, 44.84.

[000366] 3-thienyl acetaldehyde cyanohydrin:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 500 MHz)  $\delta$  7.09 (m, 3H), 4.60 (t,  $J = 6.25$  Hz, 1H), 3.12 (d,  $J = 6.25$  Hz, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 125 MHz)  $\delta$  129.05, 127.16, 125.27, 122.65, 119.87, 61.58, 44.90.

15 [000367] Preparation of racemic mandelic acids standards from corresponding cyanohydrins: (Stoughton, R.W. J. Am. Chem. Soc. 1941, 63, 2376) 2-bromomandelonitrile (230 mg, 1.08 mmol) was dissolved in conc. HCl (1 mL) and stirred at room temperature for 18 h and then at 70 °C for 24 h. After cooling, the reaction mixture was extracted with diethyl ether (4 x 2 mL). Organic extracts were, combined, dried over  $\text{MgSO}_4$ , filtered and  
20 concentrated *in vacuo*. 2-bromomandelic acid was isolated as a colorless powder (180 mg, 0.78 mmol, 70 % yield).

[000368] Preparation of racemic aryllactic acids standards from corresponding amino acids: Phenylalanine (10 mmol, 1.65g) was dissolved in 30 ml 2N  $\text{H}_2\text{SO}_4$  at room temperature under  $\text{N}_2$  (g) atmosphere. Sodium nitrite (1.4 g in 3 ml aqueous solution, 2 eq) solution was added  
25 slowly to the reaction mixture over a period of 3-4 hours with vigorous stirring at room temperature under  $\text{N}_2$  (g) atmosphere. The reaction mixture was stirred overnight and the phenyllactic acid product was then extracted into diethylether (3 x 30 ml). Combined ether extracts were dried over  $\text{MgSO}_4$  and then filtered and concentrated *in vacuo*. (Kenji, I.; Susumu, A.; Masaru, M.; Yasuyoshi, U.; Koki, Y.; Koichi, K. Patent Number, WO0155074,  
30 Publication date: 2001-08-02.)

General Method for Enzymatic Preparation of  $\alpha$ -hydroxy acids:

[000369] (*R*)-(-)-Mandelic Acid To a solution of mandelonitrile (1.005 g, 7.56 mmol) in 150 mL of sodium phosphate (100 mM) buffer at pH 8 with 10% v/v methanol, that had been



$\text{N}_2$  (g) sparged, at 37 °C, was added 9 mg of nitrilase 1 (normalized for nitrilase content). The reaction was conducted under  $\text{N}_2$  (g) atmosphere on a rotating platform shaker. Reaction progress was monitored by withdrawing aliquots for HPLC analysis. After 3 h incubation, the reaction mixture was acidified to pH 2 with 1 N HCl and extracted with diethyl ether (4 x 50 ml). Organic fractions were concentrated *in vacuo* and then the residue was taken up in 10% sodium bicarbonate solution. This aqueous solutions was then washed with diethyl ether (3 x 50 ml) and then acidified to pH 2 with 1 N HCl and extracted with diethyl ether (3 x 50 ml). Organic fractions were combined, washed with brine, dried over  $\text{MgSO}_4$ , filtered and then concentrated *in vacuo*. (R)-(-)-Mandelic acid (933 mg, 6.22 mmol) was isolated as a colorless powder in 86 % yield.  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ , 500 MHz)  $\delta$  12.6 (br, s, 1H) 7.41 (m, 2H), 7.34 (m, 2H), 7.28 (m, 1H), 5.015 (s, 1H).  $^{13}\text{C}$  NMR  $\text{DMSO-d}_6$ , 125 MHz)  $\delta$  174.083, 140.216, 128.113, 127.628, 126.628, 72.359. MS calc'd for  $[\text{C}_8\text{H}_8\text{O}_3]$  150.07, found 150.9 (ESI +); *ee* = 98 % [HPLC].  $[\alpha]^{20}_{598} = -134.6$  (*c* = 0.5, methanol).

**[000370]** (-)-2-chloromandelic acid  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ , 500 MHz)  $\delta$  7.75 (m, 1H), 7.44 (m, 1H), 7.34 (m, 2H), 5.34 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{DMSO}$ , 298K, 125MHz)  $\delta$  173.070, 137.985, 132.105, 129.399, 129.158, 128.705, 127.235. MS calc'd for  $[\text{C}_8\text{H}_7\text{ClO}_3]$  186.0, found 185.0 (LC-MS -). *ee* = 96 % [HPLC]. 92 % yield.  $[\alpha]^{20}_{598} = -137.6$  (*c* = 0.5, ethanol).

**[000371]** (-)-2-bromomandelic acid  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ , 500 MHz)  $\delta$  7.60 (d, *J* = 7.93, 1H), 7.48 (m, 1H), 7.40 (m, 1H), 7.25 (m, 1H), 5.30 (s, 1H).  $^{13}\text{C}$  NMR  $\text{DMSO-d}_6$ , 125 MHz)  $\delta$  172.994, 139.61, 132.355, 129.652, 128.753, 127.752, 122.681, 71.644. MS calc'd for  $[\text{C}_8\text{H}_7\text{BrO}_3]$  230.0, found 230.9. *ee* = 96% [HPLC]. 92% yield.  $[\alpha]^{20}_{598} = -116.4$  (*c* = 0.5, ethanol).

**[000372]** (-)-2-methylmandelic acid  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ , 500 MHz)  $\delta$  11.78 (bs, 1H) 7.38 (m, 1H), 7.16-7.38 (m, 3H), 5.18 (s, 1H), 2.35 (s, 3H).  $^{13}\text{C}$  NMR  $\text{DMSO-d}_6$ , 125 MHz)  $\delta$  174.229, 138.623, 135.649, 130.129, 127.491, 126.990, 125.698, 125.698, 69.733, 18.899. MS calc'd for  $[\text{C}_9\text{H}_{10}\text{O}_3]$  166.1, found 165.2. *ee* = 91 % [HPLC]. 86 % yield.  $[\alpha]^{20}_{598} = -164.4$  (*c* = 0.5, ethanol).

**[000373]** (-)-3-chloromandelic acid  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ , 500 MHz)  $\delta$  7.46 (s, 1H), 7.36 (m, 3H), 5.07 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{DMSO}$ , 298K, 125MHz)  $\delta$  173.554, 142.685, 132.813, 130.069, 127.568, 126.355, 125.289, 71.659. MS calc'd for  $[\text{C}_8\text{H}_7\text{ClO}_3]$  186.0, found 185.34 (MALDI TOF -). *ee* = 98 % [HPLC]. 70 % yield.  $[\alpha]^{20}_{598} = -120.4$  (*c* = 0.5, methanol).



- 5 [000374] (-)-3-bromomandelic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.60 (s, 1H), 7.49 (m, 1H), 7.42 (m, 1H), 7.31 (m, 1H), 5.06 (s, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  173.551, 142.917, 130.468, 130.379, 129.237, 125.687, 121.404, 71.605. MS calc'd for  $[\text{C}_8\text{H}_7\text{BrO}_3]$  229.98, found 229.1 (LC-MS).  $ee = 98\%$  [HPLC]. 82 % yield.  $[\alpha]^{20}_{598} = -84.8$  ( $c = 0.5$ , ethanol).
- [000375] (-)-4-fluoromandelic acid  $^1\text{H}$  NMR (DMSO, 298K, 500MHz)  $\delta$  12.65 (s, 1H), 7.44 (m, 2H), 7.17 (m, 2H), 5.91 (s, 1H), 5.03 (s, 1H)  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  173.93, 162.57, 136.47, 128.61, 128.55, 114.96, 114.80, 71.61. MS calc'd for  $[\text{C}_8\text{H}_7\text{FO}_3]$  170.0, found 168.8.  $ee = 99\%$  [HPLC]. 81% yield.  $[\alpha]^{20}_{598} = -152.8$  ( $c = 0.5$ , methanol).
- 10 [000376] (-)-1-naphthylglycolic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  8.28-8.26 (m, 1H), 7.87-7.93 (m, 2H), 7.47-7.58 (m, 4H), 5.66 (s, 1H).  $^{13}\text{C}$  NMR DMSO- $d_6$ , 125 MHz)  $\delta$  174.288, 136.284, 133.423, 130.654, 128.353, 128.192, 125.926, 125.694, 125.613, 125.266, 124.558, 70.940. MS calc'd for  $[\text{C}_{12}\text{H}_{10}\text{O}_3]$ : 202.21 found 201.37 (MALDI TOF -).  $ee = 95\%$  [HPLC]. 90 % yield  $[\alpha]^{20}_{598} = -115.4$  ( $c = 0.5$ , ethanol).
- 15 [000377] (-)-2-naphthylglycolic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  12.6 (bm, 1H), 7.88-7.93 (m, 4H), 7.48-7.56 (m, 3H), 5.20 (s, 1H).  $^{13}\text{C}$  NMR DMSO- $d_6$ , 125 MHz)  $\delta$  174.005, 137.760, 132.644, 132.498, 127.811, 127.658, 127.506, 127.209, 125.993, 125.334, 124.761, 72.472. MS calc'd for  $[\text{C}_{12}\text{H}_{10}\text{O}_3]$  202.21, found 201.37 (MALDI TOF).  $ee = 98\%$  [HPLC]. 68% yield.  $[\alpha]^{20}_{598} = -115.4$  ( $c = 0.5$ , ethanol).
- 20 [000378] (-)-3-pyridylglycolic acid This Reaction was performed in 100 mM ammonium formate buffer at pH 8. To isolate the product, the reaction mixture was filtered through a 10,000 MWCO membrane to remove enzyme and then concentrated *in vacuo*.  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  8.56 (s, 1H), 8.36 (d,  $J = 4.57$  Hz, 1H), 8.25 (s, 1H), 7.71 (m, 1H), 7.25 (dd,  $J = 4.98, 4.80$  Hz 1H), 5.45 (s, 1H).  $^{13}\text{C}$  NMR DMSO- $d_6$ , 125 MHz)  $\delta$  165.911, 147.862, 147.251, 139.118, 133.381, 122.746, 71.508. MS calc'd for  $[\text{C}_7\text{H}_7\text{NO}_3]$  153.04, found 154.0 ((MALDI TOF).  $ee = 92\%$  [HPLC], 84% yield,  $[\alpha]^{20}_{598} = -65.2$  ( $c = 0.5$ ,  $\text{H}_2\text{O}$ ).
- 25 [000379] (-)-3-thienylglycolic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.48 (m, 1H), 7.45 (d,  $J = 2.81$ , 1H), 7.10 (m, 1H), 5.09 (s, 1H), 3.33 (s, 1H)  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  173.704, 141.109, 126.446, 126.042, 122.247, 68.915 MS calc'd for  $[\text{C}_6\text{H}_6\text{O}_3\text{S}]$  158.00, found 157.224 (MALDI TOF).  $ee = 95\%$  [HPLC]. 70 % yield.  $[\alpha]^{20}_{598} = -123.2$  ( $c = 0.5$ , methanol).
- 30



- [000380] (S)-(-)-phenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.28(m, 5H), 4.17(dd,  $J = 4.5$  Hz,  $J = 8.3$  Hz, 1H), 2.98(dd,  $J = 4.5$  Hz,  $J = 13.7$  Hz, 1H), 2.79 (dd,  $J = 8.3$  Hz,  $J = 13.7$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  178.16, 133.4, 129.27, 128.6, 127.3, 70.45, 44.12.  $ee = 97\%$  [HPLC], 84 % yield.  $[\alpha]^{20}_{598} = -17.8$  ( $c = 0.5$ , methanol).
- 5 [000381] (-)-2-methylphenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.16 (m, 4H), 4.47 (dd,  $J = 3.9$  Hz,  $J = 8.8$  Hz, 1H), 3.25(dd,  $J = 3.9$ Hz, 14.3 Hz, 1H), 2.94 (dd,  $J = 8.8$  Hz,  $J = 14.3$ Hz), 2.35(s, 3H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  178.61, 137.08, 134.74, 130.80, 130.25, 127.44, 126.34, 70.93, 37.67, 19.79. MS calc'd [ $\text{C}_{10}\text{H}_{12}\text{O}_3$ ] 180.08, found 180.0 (ESI +). 86 % yield.  $ee = 95\%$  [HPLC].  $[\alpha]^{20}_{598} = -13.2$  ( $c = 0.5$ , methanol).
- 10 [000382] (-)-2-bromophenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.28 (m, 4H), 4.60(dd,  $J = 4.0$  Hz,  $J = 9.1$  Hz, 1H), 3.45(dd,  $J = 4.0$  Hz,  $J = 14.1$  Hz, 1H), 3.04(dd,  $J = 8.0$  Hz,  $J = 14.1$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  178.70, 136.05, 133.21, 132.10, 128.99, 127.72, 125.0, 70.04, 40.76. MS calc'd for [ $\text{C}_9\text{H}_9\text{BrO}_3$ ] 243.9, found 243.3 (ESI +). 91 % yield.  $ee = 93\%$  [HPLC],  $[\alpha]^{20}_{598} = -17.6$  ( $c = 0.5$ , methanol)
- 15 [000383] (-)-2-fluorophenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.10 (m, 4H), 4.64 (t,  $J = 6.8$  Hz, 1H), 3.11(d,  $J = 6.8$  Hz, 2H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  132.18, 131.52, 129.66, 129.03, 128.07, 124.05, 115.8, 63.02, 44.79. MS calc'd for [ $\text{C}_9\text{H}_8\text{FNO}$ ]: 165.06, found 164.2 (ESI +). 91 % yield.  $ee = 88\%$  [HPLC].  $[\alpha]^{20}_{598} = -14.0$  ( $c = 0.5$ , methanol).
- 20 [000384] (-)-3-methylphenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.18 (m, 1H), 7.02 (m, 3H), 4.54 (dd,  $J = 4.6$  Hz,  $J = 8.0$  Hz, 1H), 3.06(dd,  $J = 4.54$  Hz,  $J = 14.4$  Hz, 1H), 2.83(dd,  $J = 8.0$  Hz,  $J = 14.4$  Hz, 1H), 2.36 (s, 3H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  175.88, 163.80, 130.33, 130.09, 125.7, 116.68, 113.75, 71.31, 34.28. MS calc'd for [ $\text{C}_{10}\text{H}_{11}\text{NO}$ ] 161.08, found 162.2 (ESI +). 80 % yield.  $ee = 98\%$  [HPLC].  $[\alpha]^{20}_{598} = -2.4$  ( $c = 0.5$ , methanol).
- 25 [000385] (-)-3-fluorophenyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.2 (m, 1H), 6.9 (m, 3H), 4.56 (dd, 4.5 Hz,  $J = 7.9$  Hz, 1H), 3.09(dd,  $J = 4.5$  Hz,  $J = 14.1$  Hz, 1H), 2.86 (dd,  $J = 7.9$  Hz,  $J = 14.1$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  175.88, 163.80, 130.33, 130.09, 125.7, 116.68, 113.75, 71.31, 34.28. MS calc'd for [ $\text{C}_9\text{H}_9\text{O}_3\text{F}$ ] 184.05, found 184.1 (ESI +). 82 % yield.  $ee = 97\%$  [HPLC].  $[\alpha]^{20}_{598} = -5.2$  ( $c = 0.5$ , methanol).
- 30 [000386] (-)-1-naphyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  8.57 (m, 1H), 8.21(m, 1H), 8.08 (m, 1H), 7.61 (m, 4H), 4.64 (dd, 3.5 Hz, 8.5 Hz, 1H), 3.84 (dd,  $J = 3.5$  Hz,  $J = 14.5$



Hz, 1H), 3.38 (dd,  $J = 8.5$  Hz,  $J = 14.5$  Hz, 1H)  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  177.7, 140.31, 129.74, 129.24, 128.92, 128.26, 127.84, 125.63, 124.53, 124.05, 123.42, 70.58, 38.0. MS calc'd for  $[\text{C}_{13}\text{H}_{11}\text{NO}]$  197.08, found 197.1 (ESI +). 87 % yield.  $ee = 94$  % [HPLC].  $[\alpha]_{598}^{20} = -16.2$  ( $c = 0.5$ , methanol).

- 5    **[000387]**    (-)-2-pyridyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  8.49 (m, 1H), 7.62 (m, 1H), 7.21 (m, 2H), 4.50 (t,  $J = 5.0$  Hz, 1H), 3.01 (d,  $J = 5.0$  Hz, 2H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125 MHz)  $\delta$  178.8, 159.79, 148.84, 136.89, 124.35, 121.75, 71.14, 44.09. MS calc'd for  $[\text{C}_8\text{H}_9\text{NO}_3]$ : 167.06, found 167.0. (ESI +). 62 % yield.  $ee = 94$  % [HPLC],  $[\alpha]_{598}^{20} = -3.6$  ( $c = 0.5$ , methanol).
- 10   **[000388]**    (-)-3-pyridyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  8.43(m, 2H), 7.62(m, 1H), 7.28(m, 1H), 4.57(t, 5.37Hz, 1H), 2.85(d, 5.37Hz, 2H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125 MHz)  $\delta$  176.6, 150.03, 147.12, 136.41, 129.45, 123.26, 61.56, 31.46 MS calc'd for  $[\text{C}_8\text{H}_9\text{NO}_3]$  167.06, found 167.0 (ESI +). 59 % yield.  $ee = 94$  % [HPLC].  $[\alpha]_{598}^{20} = -4.0$  ( $c = 0.5$ , methanol).
- 15   **[000389]**    (-)-2-thienyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.18(m, 1H), 6.94(m, 1H), 6.90 (m, 1H), 4.49 (dd,  $J = 4.1$  Hz,  $J = 6.25$  Hz, 1H), 3.36 (dd,  $J = 4.1$  Hz,  $J = 15.0$  Hz, 1H), 3.26(dd,  $J = 6.25$  Hz,  $J = 15.0$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  127.68, 127.41, 125.58, 124.60, 118.70, 63.25, 44.84. MS calc'd for  $[\text{C}_7\text{H}_7\text{NOS}]$  153.02, found 153.0 (ESI +). 85 % yield.  $ee = 95$  % [HPLC].  $[\alpha]_{598}^{20} = -13.0$  ( $c = 0.5$ , methanol).
- 20   **[000390]**    (-)-3-thienyllactic acid  $^1\text{H}$  NMR (DMSO- $d_6$ , 500 MHz)  $\delta$  7.30(m, 1H), 7.13(m, 1H), 7.01(m, 1H), 4.50 (dd,  $J = 4.25$  Hz,  $J = 6.5$  Hz, 1H), 3.21(dd,  $J = 4.25$  Hz,  $J = 15.0$  Hz, 1H), 3.10 (dd,  $J = 6.5$  Hz,  $J = 15.0$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  127.50, 136.09, 128.83, 126.24, 123.32, 70.65, 34.84. MS calc'd for  $[\text{C}_7\text{H}_8\text{O}_3\text{S}]$  172.02, found 172.1 (ESI +). 81 % yield.  $ee = 96$  % [HPLC].  $[\alpha]_{598}^{20} = -18.8$  ( $c = 0.5$ , methanol).

25    Enzymatic Hydrolysis of 3-Hydroxyglutarylnitrile:

- [000391]**    3-Hydroxyglutarylnitrile (1.0 g, 9.0 mmol, 240 mM) was suspended in  $\text{N}_2$  (g) sparged sodium phosphate buffer (37.5 mL, pH 7, 100 mM) at room temperature. Cell lysate (30 mg, normalized for nitrilase content) was added to bring the concentration to 0.8 mg/ml enzyme and the reaction was at shaken at 100 rpm, room temperature. Reaction progress was
- 30    monitored by TLC (1:1 EtOAc:Hexanes,  $R_f = 0.32$ , nitrile;  $R_f = 0.0$ , acid) After 22 h, the reaction was acidified with 1M HCl. The reaction mixture was continuously extracted with diethyl ether. The acid product was isolated as a yellow oil (1.15 g, 98 % yield).  $^1\text{H}$  NMR



(DMSO, 298K, 500MHz)  $\delta$  12.32 (s, 1H), 5.52 (s, 1H), 4.10 (m, 1H), 2.70 (dd, 1H,  $J = 16.8, 4.1$  Hz), 2.61 (dd, 1H,  $J = 16.9, 6.3$  Hz), 2.44 (dd, 1H,  $J = 15.4, 5.3$  Hz), 2.37 (dd, 1H,  $J = 15.6, 7.8$  Hz).  $^{13}\text{C}$  NMR (DMSO, 298K, 125 MHz)  $\delta$  171.9, 118.7, 63.4, 41.2, 25.2 MS calc'd for  $[\text{C}_5\text{H}_7\text{NO}_3]$ : 129.0, found 130.0  $[\text{M}+\text{H}^+]$ , (ESI +).

#### 5 Preparation of (R)-(-)-Methyl (3-O-[benzoyl]-4-cyano)-butanoate

[000392] Benzoyl chloride (0.068 ml, 0.752 mmol) was added to a stirred solution of (*R*)-methyl-(3-hydroxy-4-cyano)-butanoate (71.7 mg, 0.501mmol) in pyridine (2.0 ml), at room temperature. After 19 hours, add an additional 0.5 equivalent of benzoyl chloride (0.023ml, 0.251mmol). Reaction was complete at 23 h, as determined by TLC. Add 1ml  $\text{H}_2\text{O}$ , extract  
 10 with ether (3 x 10ml). Wash with brine (2 x 10ml). Dry combined aqueous extracts with  $\text{MgSO}_4$ . Filter off drying agent and remove solvent by rotary evaporation. Purify by column chromatography (hexane:ethyl acetate [2:1]. Rotary evaporation of fractions yielded the product as a yellow oil (46 mg, 0.186 mmol, 37%).  $^1\text{H}$  NMR (DMSO, 298K, 500MHz)  $\delta$  7.96 (d, 2H,  $J = 7.8$ ), 7.70 (t, 1H,  $J = 7.25$ ), 7.56 (t, 2H,  $J = 7.8$ ), 5.55 (m, 1H), 3.59 (s, 3H),  
 15 3.13 (m, 2H), 2.90 (m, 2H).  $^{13}\text{C}$  NMR (DMSO, 298K, 125MHz)  $\delta$  169.6, 164.5, 133.8, 129.3, 128.9, 128.5, 117.3, 66.0, 51.8, 37.5, 22.2 MS calc'd for  $[\text{C}_{13}\text{H}_{13}\text{NO}_4]$ : 247.25, found 270.3  $[\text{M}+\text{Na}^+]$   $ee = 95\%$  [HPLC].  $[\alpha]^{20}_{598} -32.4$  ( $c = 0.5$ ,  $\text{CHCl}_3$ ).

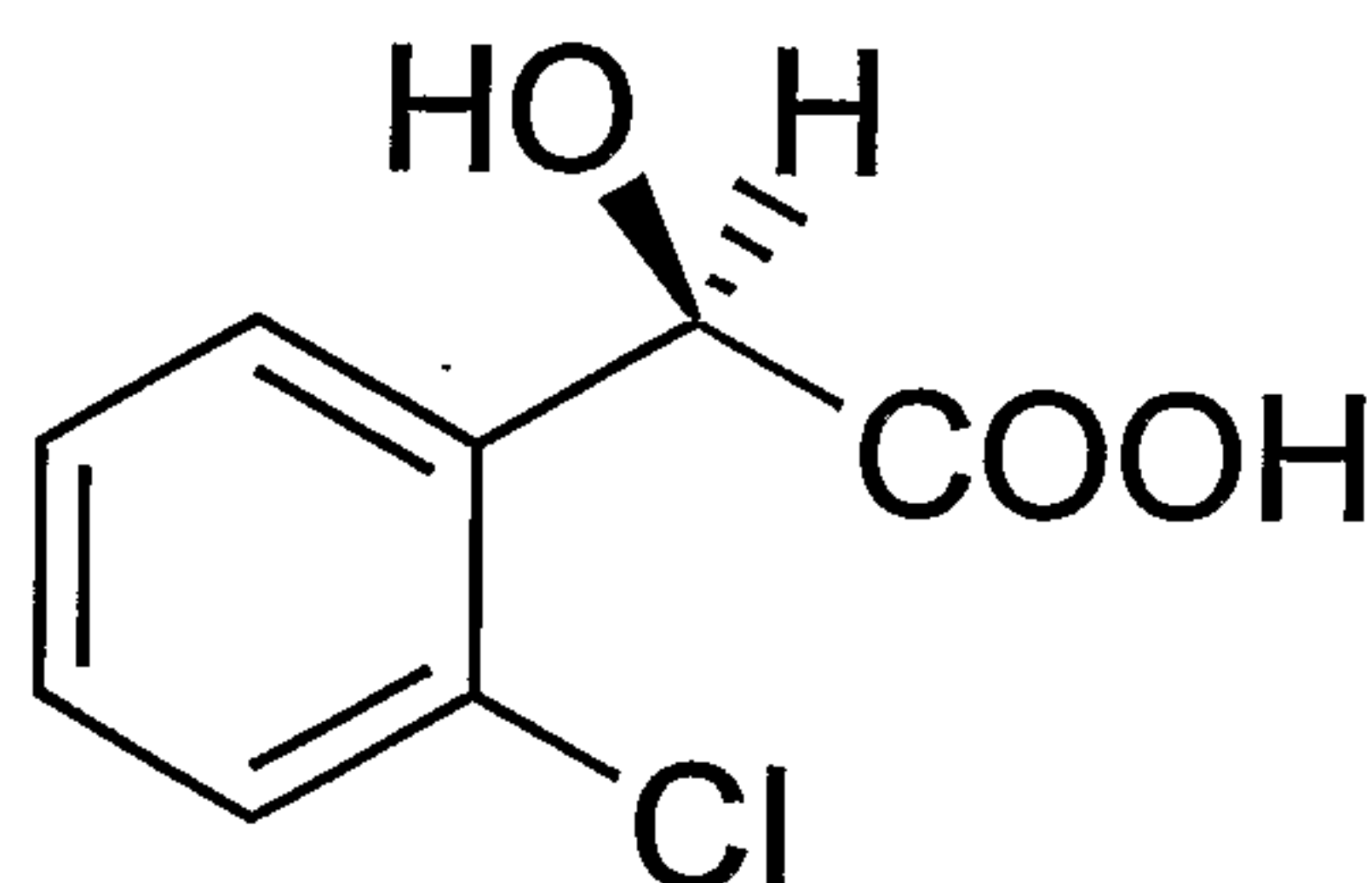
#### Synthesis of (R)-Ethyl-(3-hydroxy-4cyano)-butanoate

[000393] A 0.2 M solution of (*R*)-3-hydroxy-4-cyano-butanoic acid (50 mg, 0.387 mmol) in  
 20 anhydrous ethanol (1.94 mL) was prepared. The ethanol solution was added dropwise to 1.0 ml of a 50:50 (v/v) mixture of anhydrous 1 M HCl ethereal solution and anhydrous ethanol over sieves. The reaction was stirred overnight at room temperature under  $\text{N}_2$  (g) atmosphere. The reaction was monitored by TLC, (1:1 EtOAc:Hexanes,  $R_f = 0.45$ , ester;  $R_f = 0.0$ , acid, stained with p-anisaldehyde). After 30 hrs, solvent was removed by rotary evaporation. The  
 25 crude product was taken up in 25 mL ether, washed with 5 mL saturated bicarbonate and then 5 mL brine. The organic extract was dried over  $\text{MgSO}_4$ , filtered and then concentrated *in vacuo*, yielding the product as a clear oil.  $^1\text{H}$  NMR (DMSO, 298K, 500MHz)  $\delta$  5.60 (d, 1H,  $J = 5.58$  Hz), 4.12 (m, 1H), 4.07 (q, 2H,  $J = 7.1$ ), 2.66 (m, 2H), 2.47 (m, 2H), 1.87 (t, 3H,  $J = 7.0$ ).  $^{13}\text{C}$  NMR (DMSO, 298K, 125 MHz)  $\delta$  170.21, 118.60, 63.40, 59.98, 41.10, 25.14,  
 30 14.02. MS calc'd for  $[\text{C}_7\text{H}_{11}\text{NO}_3]$ : 157.1, found 158.2.  $[\text{M}+\text{H}^+]$



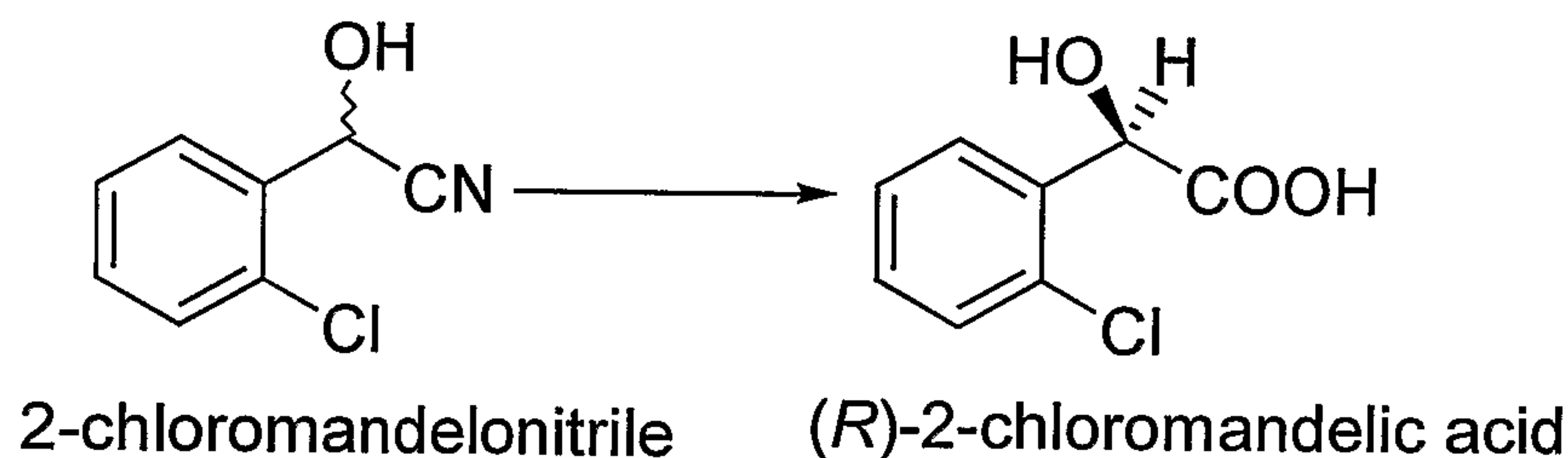
Example 13: Optimization Of Nitrilases For The Enantioselective Production Of  
(R)-2-Chloromandelic Acid

[000394] Chloromandelic acid has the structure:



5

[000395] Nitrilases were identified which selectively produced (*R*)-2-chloromandelic acid from (*R,S*)-2-chloromandelonitrile. Nitrilases were identified which were useful to improve the enantioselectivity of the enzymes and establishing the effects of process conditions on the enzymes. An examination of the reaction conditions for the enzymatic nitrile hydrolysis was carried out in order to improve the enantiomeric excess of the product. Additionally, further investigation into the effects of process conditions on the enzyme was performed.



15 [000396] In this aspect, the enantioselective production of (*R*)-2-chloromandelic acid was the target. One enzyme, SEQ ID NOS:385, 386, was selected for further confirmation of its enantioselectivity on 2-chloromandelonitrile. SEQ ID NOS:385, 386 was shown to be stable to process components, with a half-life of 8 hours. The enzyme was inhibited by 2-chlorobenzaldehyde and a contaminant in the cyanohydrin substrate, 2-chlorobenzoic acid.

20 The enzymatic reaction was scaled up to a substrate concentration of 45 mM 2-chloromandelonitrile. Over 90% conversion was obtained, with *ee* of 97%. The chiral HPLC method was improved, to remove a contaminating peak that was present in the substrate. Improved accuracy in the determination of enantioselectivity was obtained using this method.



[000397] Nitrilases were screened against 2-chloromandelonitrile, with 31 nitrilases exhibiting activity on this substrate. High enantioselectivities were shown by 9 enzymes. The optimization of 5 of these enzymes was undertaken and one of them was identified as a candidate for the next stage of development.

5 [000398] In an effort to improve the enantioselectivity of the selected enzymes for (*R*)-2-chloromandelic acid, a number of factors that are known to affect this property, together with the activity of the enzymes, were investigated. These included pH, temperature, buffer strength and addition of solvents to the reaction. Initially, 5 nitrilases were selected for these studies, based on the high enantioselectivities obtained by these enzymes. These enzymes  
10 were: SEQ ID NOS:385, 386, SEQ ID NOS:197, 198, SEQ ID NOS:217, 218, SEQ ID NOS:55, 56, and SEQ ID NOS:167, 168.

#### Effect of pH

[000399] The enzymatic reactions were run at a range of pH values, from pH 5 to pH 9. An increase in both activity and enantioselectivity with increasing pH was observed for all of  
15 the enzymes. With the exception of SEQ ID NOS:385, 386, pH 9 (0.1 M Tris-HCl buffer) was determined as the optimum for activity and enantioselectivity. The optimum pH for SEQ ID NOS:385, 386 was pH 8 (0.1 M sodium phosphate buffer).

#### Effect of temperature

[000400] The enzymes exhibited similar temperature profiles, with the highest activities  
20 being measured at 37°C and 45°C. Although the latter temperature resulted in higher conversions, the enantioselectivity of most of the enzymes showed a clear preference for the lower temperatures, with ee values being 10-20% lower when the temperature was raised above 37°C. In the case of SEQ ID NOS:385, 386 a slight optimum for enantioselectivity was evident at 37°C. Therefore, this temperature was established as the optimum for  
25 hydrolysis of 2-chloromandelonitrile by these enzymes.

#### Effect of enzyme concentration

[000401] During the concurrent investigation into the enantioselective hydrolysis of phenylacetaldehyde cyanohydrin to L-phenyllactic acid, the concentration of the enzyme in the reaction was found to have a significant effect on the enantioselectivity of the reaction.  
30 This provided an indication that the enzymatic hydrolysis rate was faster than the rate of racemization of the remaining cyanohydrin in the reaction. On this basis, the effect of enzyme concentration on the enantioselectivity of the enzymes towards (*R*)-2-chloromandelonitrile was investigated. Enzymatic reactions were performed with the



standard concentration of enzyme (0.6 mg protein/ml), half the standard concentration and one-tenth of the standard concentration.

[000402] The following Table indicates the highest conversions achieved for the reactions, with the corresponding ee. With the exception of SEQ ID NOS:385, 386, it

5 appears that very little, if any, increased enantioselectivity is observed. Therefore, it appears that the rate of racemization of the remaining chloromandelonitrile is not a limiting factor to obtaining higher enantioselectivities.

#### Investigation of other positive enzymes

[000403] In addition to the enzymes in the above Table, a number of other nitrilases  
10 were screened for their enantioselectivities on 2-chloromandelonitrile. Some of these enzymes were newly discovered enzymes. Some were reinvestigated under conditions that have since been found to be optimal for these enzymes (pH 8 and 37°C). The results of this screening are shown below in the Table.

#### Effect of co-solvent concentration

15 [000404] The addition of methanol as a cosolvent in the enzymatic reactions was shown to enhance the ee. In order to establish the lowest level of methanol that could be added to the reactions, the enzyme reactions were performed at varying concentrations of methanol, ranging from 0-20% (v/v). No significant differences in enantioselectivity were evident  
20 between the various methanol concentrations. However, the ee in these reactions was 97-98%, while that of the control reaction, with no added methanol was 95-96%. While this difference in ee is small, the effect of the methanol was shown in more than one set of experiments during the course of this investigation and is therefore regarded as significant.

#### Effect of reaction components on activity of SEQ ID NOS:385, 386

25 [000405] A vital part of an investigation into process optimization of an enzyme involves the determination of the effects of any compounds which could be present in the enzymatic reaction. For SEQ ID NOS:385, 386, these components were established as the starting material and equilibrium product of the cyanohydrin, 2-chlorobenzaldehyde; the product, 2-chloromandelic acid and the contaminant detected in the substrate, 2-  
30 chlorobenzoic acid. The addition of cyanide to the reaction was found to have no effect on the enzyme activity. The presence of trace amounts of triethylamine was also found to be tolerable to the enzyme.



[000406] The effect of the various reaction components on the activity of SEQ ID NOS:385, 386 was assessed by addition of various levels of possible inhibitors to the enzyme reaction. From these experiments, it appeared that both the aldehyde and its oxidation product, 2-chlorobenzoic acid were detrimental to enzyme activity. Approximately 70% and 40% of the activity of SEQ ID NOS:385, 386 was lost upon addition of 5 mM 2-chlorobenzaldehyde or 5 mM 2-chlorobenzoic acid to the reaction, respectively.

#### Scale-up hydrolysis of 2-chloromandelonitrile

[000407] In order to confirm the conversion and enantioselectivity obtained by SEQ ID NOS:385, 386 for the production of (R)-2-chloromandelic acid, a larger scale reaction was performed and the product isolated from the aqueous mixture. The reaction was performed in a 20 ml reaction volume, with a substrate concentration of 45 mM 2-chloromandelonitrile. Complete conversion of the cyanohydrin was obtained, with 30 mM product formed. The ee of the product was 97% and the specific activity of the enzyme was 0.13 mmol product/mg nitrilase/h.

[000408] It is evident from this experiment, together with the other experiments performed, that the formation of product does not account for the complete loss of substrate. In all experiments, a nitrile-containing control sample was run, in order to determine the extent of breakdown of the cyanohydrin. Overall, it appears that approximately 50% of the substrate is lost over a period of 4 hours at 37°C. It is expected that this breakdown would be to its equilibrium products, cyanide and 2-chlorobenzaldehyde, which could undergo further oxidation. A larger scale reaction was also run at a substrate concentration of 90 mM 2-chloromandelonitrile. However, no product was detected in this reaction. At higher substrate concentrations, it is expected that the concentration of the equilibrium product, 2-chlorobenzaldehyde and the contaminant, 2-chlorobenzoic acid will be present in higher amounts. Based on the results above, it is possible that the enzyme will be completely inhibited under such conditions.

#### Reactions under biphasic conditions

[000409] The use of biphasic systems can facilitate product recovery following the enzymatic reaction step. These systems can be also be used for the removal of products or by-products which are inhibitory to the enzyme. The nitrilases were shown to be active under biphasic conditions using a variety of solvents. Following the low conversions obtained at the higher substrate concentration above, further investigation of a biphasic system was performed with the hit enzyme, SEQ ID NOS:385, 386. It was important to ascertain



whether any inhibitory factors could be removed by the solvent phase and whether any process advantages could be gained by the use of a biphasic system.

[000410] Promising results were obtained with hexane as the organic phase. Therefore, further investigations involved the use of this solvent at two different levels: 100% and 70% of the volume of the aqueous phase, with increasing substrate concentrations, up to 90 mM. The substrate was dissolved in the organic phase. The level of hexane did not appear to affect the level of product formation, particularly at the higher concentrations of 2-chloromandelonitrile.

[000411] Once again, high conversion was observed in a biphasic system, with a 76% yield of product being observed after 5 hours. The rate of product formation appeared to be slightly lower than in the corresponding monophasic system, where the reaction is complete within 1 hour. Lower enantioselectivity was observed in the biphasic system. Some possibilities which may account for these results are (i) the mass transfer rate is lower than the rate of enzyme activity or (ii) the non-polar solvent directly affects the enzyme.

[000412] At a higher substrate concentration, a very low conversion was observed, with 7 mM 2-chloromandelic acid being formed from 90 mM 2-chloromandelonitrile. This level of conversion, albeit low, was higher than that observed in the monophasic system with the same substrate concentration. These results suggest that some of the inhibitory 2-chlorobenzaldehyde or 2-chlorobenzoic acid is retained in the non-polar organic solvent.

#### Standard assay conditions:

[000413] The following solutions were prepared:

- Substrate stock solution: 50 mM of the cyanohydrin substrate in 0.1 M phosphate buffer (pH 8).
- Enzyme stock solution: 3.33 ml of 0.1 M phosphate buffer (pH 8) to each vial of 20 mg of lyophilized cell lysate (final concentration 6 mg protein/ml)

[000414] The reaction volumes varied between the different experiments, depending on the number of time points taken. Unless otherwise noted, all reactions consisted of 25 mM 2-chloromandelonitrile and 10% (v/v) of the enzyme stock solution (final concentration 0.6 mg protein/ml). The reactions were run at 37°C, unless otherwise stated. Controls to monitor the nitrile degradation were run with every experiment. These consisted of 25 mM 2-chloromandelonitrile in 0.1 M phosphate buffer (pH 8).

[000415] Sampling of reactions: The reactions were sampled by removing an aliquot from each reaction and diluting these samples by a factor of 8. Duplicate samples were taken for



analysis by chiral and achiral HPLC methods. The reactions were sampled at 0.5, 1, 1.5, 2, 3, and 4 hours, unless otherwise shown in the figures above.

#### HPLC methods

The achiral HPLC method was run on a SYNERGI-RP™ column (4 μm; 50 x 2 mm) with a mobile phase of 10 mM Na phosphate buffer (pH 2.5). A gradient of methanol was introduced at 3.5 min and increased to 50% over 1.5 min, following which the methanol was decreased to 0%. Elution times for 2-chloromandelic acid and 2-chloromandelonitrile were 2.5 and 6.1 minutes, with another peak appearing with the nitrile at 5.9 minutes.

[000416] As described above, the chiral HPLC method was optimized during the course of the investigation, to improve the separation between 2-chlorobenzoic acid and (S)-2-chloromandelic acid. The optimized method was used during the latter half of the investigation and was run on a CHIROBIOTIC-R™ column. The mobile phase was 80% Acetonitrile:20% of 0.5% (v/v) acetic acid. Elution times for (S)-2-chloromandelic acid and (R)-2-chloromandelic acid were 2.4 and 3.5 minutes respectively. A peak for 2-chlorobenzoic acid eluted at 1.9 minutes. For each experiment, a standard curve of the product was included in the HPLC run. The concentration of product in the samples was calculated from the slope of these curves.

#### Effect of pH

[000417] The effect of pH on the enzyme activity and enantioselectivity was studied by performance of the standard assay in a range of different buffers: 0.1 M Citrate Phosphate pH 5; 0.1 M Citrate Phosphate pH 6; 0.1 M Sodium Phosphate pH 6; 0.1 M Sodium Phosphate pH 7; 0.1 M Sodium Phosphate pH 8; 0.1 M Tris-HCl pH 8; and 0.1 M Tris-HCl pH 9. The standard enzyme concentration was used for all enzymes, with the exception of SEQ ID NOS:385, 386, where half the standard concentration was used (5% v/v of the enzyme stock solution).

#### Effect of temperature

[000418] The effect of temperature on the activity and enantioselectivity was investigated by performing the standard assay at a range of different temperatures: room temperature, 37°C, 45°C, 50°C and 60°C. The standard enzyme concentration was used for all enzymes, with the exception of SEQ ID NOS:385, 386, where half the standard concentration was used (5% v/v of the enzyme stock solution).



Effect of enzyme concentration

[000419] Reactions were run under standard conditions, with varying enzyme concentrations: 1%, 5% and 10% (v/v) of the enzyme stock solution. The reaction volume was normalized with the appropriate buffer.

5 Addition of solvents

[000420] The enzyme reactions were performed in the presence of methanol as a cosolvent. Methanol was added to the standard reaction mixture at the following levels: 0, 5, 10, 15 and 20% (v/v).

10 [000421] Biphasic reactions with hexane were also investigated. The aqueous phase contained 10% (v/v) of the enzyme stock solution in 0.1 M phosphate buffer (pH 8). The cyanohydrin was dissolved in the hexane, prior to addition to the reaction. Two levels of organic phase were used: 1 equivalent and 0.7 equivalents of the aqueous phase volume. In addition, a range of nitrile concentrations was investigated: 25, 45 and 90 mM. These reactions were run at room temperature.

15 [000422] Samples from these reactions were taken both from the aqueous and the solvent phase. The hexane was evaporated by centrifugation under vacuum and redissolved in a 50:50 mixture of methanol and water, so that the samples were at the same dilution as the aqueous samples. Analysis of the samples was performed by non-chiral and chiral HPLC.

Effect of process components

20 [000423] (i) Activity: The effect of the process components on the activity of the enzymes was established by addition of the individual components, 2-chlorobenzaldehyde, 2-chlorobenzoic acid or 2-chloromandelic acid, to the enzymatic reaction. The enzymatic reactions were carried out under standard conditions, in the presence of one of the 2 possible inhibitors as follows: 5, 10, 20 and 25 mM 2-chlorobenzaldehyde; 1.5 and 5 mM 2-  
25 chlorobenzoic acid; and 10, 20, 40 and 80 mM 2-chloromandelic acid. Control reactions were performed under standard conditions, with no additive. At each of the sampling times, the samples were diluted to a level of 1 in 10. Control samples containing the reaction components without enzyme were used and diluted to the same level. The samples were analysed by non-chiral HPLC.

30 [000424] (ii) Stability: The stability of the enzymes to process conditions was monitored by incubation of the enzymes in the presence of the reaction components, 2-chlorobenzaldehyde and 2-chloromandelic acid for predetermined time periods, prior to assay of the enzyme activity under standard conditions. In these experiments, the enzymes were



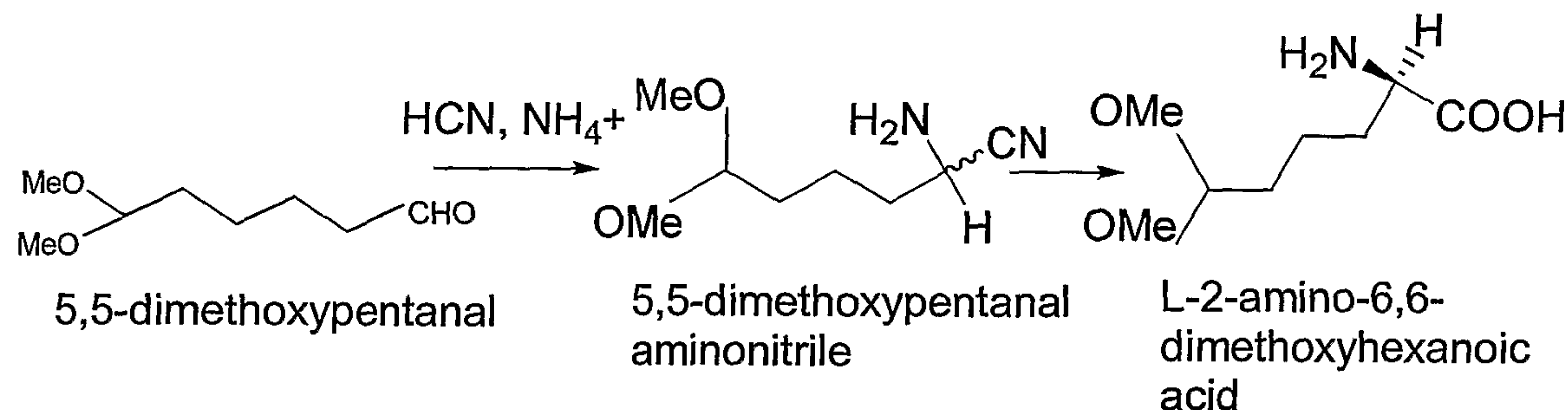
incubated at a concentration of 3 mg protein/ml in the presence of each of the following reaction components: 5, 10, 20 and 25 mM 2-chlorobenzaldehyde; and 10, 20, 40 and 80 mM 2-chloromandelic acid. Control reactions were performed by incubation of the enzyme in buffer only.

5 [000425] Assay conditions: At 0, 4, 8 and 24 hours of incubation in the particular additive, 20  $\mu$ l of the enzyme solution was removed and added to 60  $\mu$ l of a 41.6 mM substrate stock solution and 20  $\mu$ l buffer. The enzyme activity was thus assayed under standard conditions. The reactions were sampled 90 minutes after substrate addition and analyzed using the non-chiral HPLC method.

#### 10 Scale-up of enzymatic reaction

[000426] The enzymatic reactions were run at two difference concentrations: 45 mM and 90 mM substrate. The reactions were run under standard conditions, i.e. pH 8 (0.1 M sodium phosphate buffer), 37°C and 10% (v/v) of the enzyme stock solution. The substrate was dissolved in 10% (v/v) methanol prior to addition of the buffer. The final reaction  
15 volume was 20 ml and the reactions were performed with magnetic stirring.

#### Example 14: Optimization Of Nitrilases For The Enantioselective Production Of L-2-amino-6,6-dimethoxyhexanoic acid



[000427] Four of the isolated enzymes were shown to hydrolyze 2-amino-6-hydroxy hexanenitrile to (S)-2-amino-6-hydroxy hexanoic acid, with selectivity towards the L-enantiomer. A new target, with a similar structure to (S)-2-amino-6-hydroxy hexanoic acid was identified. A panel of the isolated nitrilases are screened against the target, 5,5-dimethoxypentanal aminonitrile. The positive enzymes are characterized on this substrate. Laboratory evolution techniques can be used to optimize these nitrilases for improved  
25 enantiospecificity towards the specified target. A primary screen is used to identify putative up-mutants, which is confirmed using HPLC.

[000428] Optimization of enzymes: GSSM™ and GeneReassembly™ can be performed on selected nitrilases, in order to improve the enantioselectivity and activity of the enzymes



for the production of L-2-amino-6,6-dimethoxyhexanoic acid. Four enzymes were identified that can hydrolyze enantioselectively 2-amino-6-hydroxy hexanenitrile to L-(S)-2-amino-6-hydroxy hexanoic acid. However, a slight structural difference is present in the new target molecule, L-2-amino-6,6-dimethoxyhexanoic acid. In order to determine whether this  
5 difference affects the activity and enantioselectivity of the enzymes, the complete spectrum of nitrilases is screened against the new target.

[000429] An enzyme exhibiting the highest combination of activity and enantioselectivity for the production of L-2-amino-6,6-dimethoxyhexanoic is selected for GSSM™. Following the mutation of the target enzyme, the resulting mutants will be  
10 screened on 5,5-dimethoxypentanal aminonitrile, using high throughput screening technology. Following confirmation of the up-mutants by HPLC analysis, the individual up-mutants will be combined in order to further enhance the properties of the mutant enzymes.

[000430] In parallel to GSSM™, a GeneReassembly™ can be performed on a combination of parent enzymes, at least one of which can be selected for activity and  
15 enantioselectivity on L-2-amino-6,6-dimethoxyhexanoic acid. At least two other nitrilases, with a high degree of homology, can be reassembled with the former enzyme(s); these enzymes will be selected in order to provide diversity to the reassembled sequences.

[000431] Crucial to the success of this evolution effort is the development of a high throughput assay for enantioselectivity. Such an assay is a novel enzyme-based  
20 enantioselectivity assay that allows for the screening of >30,000 mutants in a significantly shorter time period than the traditionally used method of HPLC.

[000432] In one aspect, a non-stochastic method, termed synthetic ligation reassembly, that is related to stochastic shuffling, except that the nucleic acid building blocks are not shuffled or concatenated or chimerized randomly, but rather are assembled non-  
25 stochastically, can be used to create variants. This method does not require the presence of high homology between nucleic acids to be shuffled. The ligation reassembly method can be used to non-stochastically generate libraries (or sets) of progeny molecules having at least  $10^{100}$  or at least  $10^{1000}$  different chimeras. The ligation reassembly method provides a non-stochastic method of producing a set of finalized chimeric nucleic acids that have an overall  
30 assembly order that is chosen by design, which method is comprised of the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, as assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.



[000433] The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, in one aspect, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends and, if more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). In a one aspect of the invention, the annealed building pieces are treated with an enzyme, such as a ligase (*e.g.*, T4 DNA ligase) to achieve covalent bonding of the building pieces.

[000434] In a another aspect, the design of nucleic acid building blocks is obtained upon analysis of the sequences of a set of progenitor nucleic acid templates that serve as a basis for producing a progeny set of finalized chimeric nucleic acid molecules. These progenitor nucleic acid templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, *i.e.* chimerized, recombined or shuffled.

[000435] In one exemplification, the invention provides for the chimerization of a family of related genes and their encoded family of related products. In a particular exemplification, the encoded products are nitrilase enzymes. Nucleic acids encoding the nitrilases of the invention can be mutagenized in accordance with the methods described herein.

[000436] Thus, according to one aspect of the invention, the sequences of a plurality of progenitor nucleic acid templates encoding nitrilases are aligned in order to select one or more demarcation points, which demarcation points can be located at an area of homology. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks to be generated. Thus, the demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the progeny molecules.

[000437] Typically a serviceable demarcation point is an area of homology (comprised of at least one homologous nucleotide base) shared by at least two progenitor templates, but the demarcation point can be an area of homology that is shared by at least half of the progenitor templates, at least two thirds of the progenitor templates, at least three fourths of the progenitor templates, and preferably at almost all of the progenitor templates. Even more preferably still a serviceable demarcation point is an area of homology that is shared by all of the progenitor templates.



[000438] In a one aspect, the ligation reassembly process is performed exhaustively in order to generate an exhaustive library. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, the assembly order (*i.e.*, the order of assembly of each building block in the 5' to 3' sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic, non-random). Because of the non-stochastic nature of the method, the possibility of unwanted side products is greatly reduced.

[000439] In another aspect, the method provides that, the ligation reassembly process is performed systematically, for example in order to generate a systematically compartmentalized library, with compartments that can be screened systematically, *e.g.*, one by one. Each compartment (or portion) holds chimeras or recombinants with known characteristics. In other words the invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, it allows a potentially very large number of progeny molecules to be examined systematically in smaller groups.

[000440] Because of its ability to perform chimerizations in a manner that is highly flexible, yet exhaustive and systematic, particularly when there is a low level of homology among the progenitor molecules, the invention described herein provides for the generation of a library (or set) comprised of a large number of progeny molecules. Because of the non-stochastic nature of the ligation reassembly method, the progeny molecules generated preferably comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. In a particularly aspect, such a generated library is comprised of greater than  $10^3$  to greater than  $10^{1000}$  different progeny molecular species.

[000441] In another exemplification, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (*e.g.*, one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that can later be optionally removed in an *in vitro* process (*e.g.*, by mutagenesis) or in an *in vivo* process (*e.g.*, by utilizing the gene splicing ability of a host organism). It is appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.



[000442] The synthetic ligation reassembly method of the invention utilizes a plurality of nucleic acid building blocks, each of which preferably has two ligatable ends. The two ligatable ends on each nucleic acid building block may be two blunt ends (i.e. each having an overhang of zero nucleotides), or preferably one blunt end and one overhang, or more preferably still two overhangs. On a double-stranded nucleic acid, a useful overhang can be a 3' overhang, or a 5' overhang. A nucleic acid building block can have a 3' overhang, a 5' overhang, two 3' overhangs, or two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design (*e.g.*, by designing sticky ends between building block nucleic acids based on the sequence of the 5' and 3' overhangs) and is not random.

[000443] According to one preferred aspect, a nucleic acid building block is generated by chemical synthesis of two single-stranded nucleic acids (also referred to as single-stranded oligos) and contacting them together under hybridization conditions so as to allow them to anneal to form a double-stranded nucleic acid building block. A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large. Preferred sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other preferred size ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between), and upper limits of from 2 bp to 100,000 bp (including every integer value in between).

[000444] According to one aspect, a double-stranded nucleic acid building block is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another aspect, the two strands of a double-stranded nucleic acid building block are complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this aspect, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. Preferably the codon degeneracy is introduced using the site-saturation mutagenesis described herein, using one or more N,N,GIT cassettes or alternatively using one or more N,N,N cassettes.

#### Example 15: Assays for Evaluation of Nitrilase Activity and Enantioselectivity



[000445] An assay method amenable to high throughput automation to increase the screening throughput both of the discovery and evolution efforts for nitrilases is described. The ideal assay is one that permits quantification of both product formation or substrate conversion and also enantiomeric excess. Two achiral and two chiral colorimetric assays that  
5 are amenable to high throughput screening were developed.

*Achiral Colorimetric Assays Developed:*

[000446] OPA assay for residual substrate. The OPA assay is Applicable to  $\alpha$ -amino or  $\alpha$ -hydroxy nitrile substrates. The lysis of whole cells is not necessary. These results were corroborated by HPLC for 2-chloromandelonitrile and phenyl acetaldehyde cyanohydrin.

10 The assay works best with aromatic nitriles. Aliphatic compounds exhibit a linear standard curve, fluorescence is reduced, reducing the efficacy of the assay.

[000447] LDH Assay for quantification and *ee* determination of hydroxyacid formed.

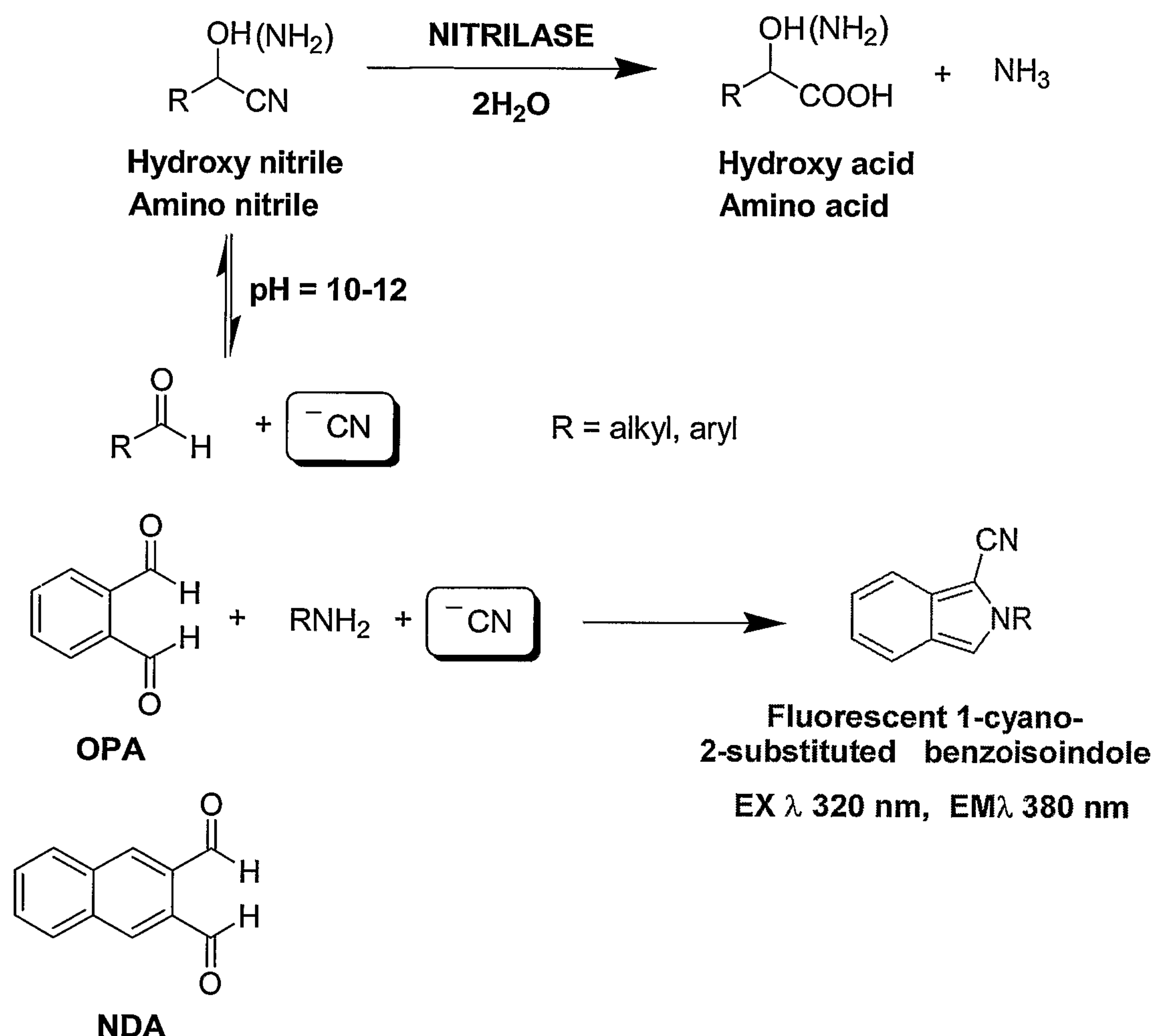
The LDH assay is applicable to phenyl lactic acid but not to 2-chloromandelic acid. Use of a resazurin detection system increases sensitivity and reduces background. Background  
15 fluorescence of whole cells was overcome either by centrifugation or heat inactivation prior to performing assay.

[000448] AAO Assay for quantification and *ee* determination of aminoacid formed. The AAO assay is applicable to phenylalanine and (S)-2-amino-6-hydroxy hexanoic acid. The use of the Amplex Red detection system increases sensitivity. Cell lysis was shown not be  
20 necessary. Cells are grown in defined media in order to prevent background fluorescence.

OPA Assay

[000449] The *o*-phthalaldehyde (OPA) fluorescence based nitrilase assay is used to quantify the amount of  $\alpha$ -hydroxynitrile substrate remaining. OPA reacts with the cyanide released from the pH controlled decomposition of  $\alpha$ -hydroxynitriles to the corresponding  
25 aldehyde and cyanide to yield a fluorescent, quantifiable product. OPA reacts with the cyanide released from the pH controlled decomposition of  $\alpha$ -hydroxynitriles to the corresponding aldehyde and cyanide to yield the fluorescent 1-cyano-2-R benzoisindole.





[000450] Standard curves were established for the following substrates: 2-Chloromandelonitrile (CMN, 0.998), Cyclohexylmandelonitrile (CHMN, 0.99), Acetophenone aminonitrile (APA, 0.99), and Phenylacetaldehyde cyanohydrin (PAC, 0.97), (Figure 5), ( $R^2$  values in parentheses). A standard curve for Phenylglycine (PGN, 0.93) was also established. Three of the substrates tested, Dimethylbutanal aminonitrile (DMB) (2-amino-4,4-dimethyl pentanenitrile), Hydroxypivaldehyde aminonitrile (HPA) and Pivaldehyde aminonitrile (PAH), gave very low fluorescence readings and unreliable results under the original assay conditions. For these compounds a number of parameters were adjusted, however the fluorescent signal strength of these compounds was not increased by these manipulations.

[000451] In an attempt to increase the fluorescent signal of these three compounds, naphthalene dicarboxaldehyde (NDA) was substituted for OPA. Standard curves for PAH, HPA and DMB with either OPA or NDA were constructed. To determine sensitivity and background fluorescence, a lyophilized nitrilase lysate (SEQ ID NOS:189, 190) with



suspected catalytic activity on each of the substrates was added. Hydrolysis was detected in three out of four of the compounds. NDA sharply boosted the signal, often by an order of magnitude, though this reduced linearity is presumably due to signal saturation.

[000452] NDA was established as an alternative detection reagent for the aliphatic compounds. However, it is desirable for the assay to utilize the same detection system for all of the substrates since this would facilitate the automated evaluation of multiple nitrilase substrates. The current OPA based assay is effective for the analysis of PAC, CMN, CHMN, APA, MN and PGN. While standard curves have been developed for the aliphatic compounds PAH, HPA, and DMB.

#### 10 Whole cell optimization

[000453] The effect of addition of lyophilized nitrilase lysate to the assay components, either untreated or heat inactivated, was evaluated. Interfering background fluorescence was not observed in either case. The OPA assay was next evaluated and optimized for nitrilase activity detection in a whole cell format. Both nitrilase expressing whole cells and *in-situ* lysed cells were evaluated. Lyophilized cell lysates were evaluated alongside their respective whole cell clones as controls. For this optimization study, mandelonitrile (MN) was chosen as a model substrate.

[000454] The lyophilized cell lysate of SEQ ID NOS:187, 188 was evaluated alongside whole cells expressing SEQ ID NOS:187, 188 and *in situ lysed* cells expressing SEQ ID NOS:187, 188. The addition of whole cells did not affect fluorescence nor result in fluorescence quenching. Addition of any of the three cell lysis solutions improved permeability (and therefore conversion) of mandelonitrile in the whole cell systems. Three cell lysing solutions were evaluated: B-PER (Pierce), BugBuster (Novagen) and CellLytic B-II (Sigma) and were found not to have a deleterious affect on the OPA assay. The addition of product  $\alpha$ -hydroxyacid or  $\alpha$ -aminoacid did not affect detection by the OPA assay.

[000455] The assay was modified from its original format, which required several liquid transfer steps, into a one plate process, where cell growth, nitrile hydrolysis and OPA assay reaction occurred in the same microtiter plate. Mandelonitrile was tested using this single well format. In this case, the *E. coli*. Gene site-saturation mutagenesis (GSSM<sup>TM</sup>) cell host was evaluated. Three clones were tested: SEQ ID NOS:101, 102, SEQ ID NOS:187, 188, and an empty vector, which was used as a control. Hydrolysis was evaluated at four timepoints, at 10 and 20 mM, and also with a 0 mM control. In an earlier experiment, clone



SEQ ID NOS:187, 188 was evaluated against the phenylacetaldehyde cyanohydrin substrate (for which this enzyme does not exhibit activity), and no activity was observed.

[000456] The OPA assay was found to detect the presence of both  $\alpha$ -hydroxy and  $\alpha$ -amino nitrile substrate. Aromatic compounds were readily detectable with the assay, while aliphatic compounds posed some detection challenges. No background issues were evident when using lyophilized cell lysates, *in-situ* lysed whole cells or unlysed whole cells. The assay is amenable to one-plate analysis, where cells are grown, incubated with the substrate, and assayed on the same plate: no liquid transfers are required, easing automation. While all nitriles tested produced a linear response, aliphatic compounds gave a low fluorescent response.

#### *Chiral LDH Assay*

[000457] A spectroscopic system based on lactate dehydrogenase (L-LDH) was developed for the analysis of the chiral  $\alpha$ -hydroxy acids which are generated by the nitrilase catalyzed hydrolysis of cyanohydrins. The hydroxynitrile substrate is not metabolized by the secondary or detection enzyme and thus starting material does not interfere. Cell lysate which is not heat treated results in background activity for the LDH system; however, heat inactivation or pelleting of the cell lysates eliminates the background activity. (See Figure 4.)

[000458] The activity and enantiomeric specificity of commercially available D- and L-lactate dehydrogenases against the nitrilases disclosed herein was evaluated. An LDH was identified which is suitable to both D- and L-phenyl lactic acid analysis. An enzyme suitable for 2-Chloromandelic acid analysis was not found. The chosen LDH enzymes exhibited virtually absolute stereoselectivity. The viability of the assay to detect D- and L-LDH produced from PAC using lyophilized cell lysate was established.

[000459] Originally, three colorimetric dyes were evaluated, all of which are tetrazolium salts: NBT (3,3'-dimethoxy-4,4'-biphenylene)bis[2,(4-nitrophenyl)-5-phenyl-2H]-, chloride) MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) INT (2-(4-Iodophenyl)-3-(4-nitrophenyl)-5-phenyl-2H-tetrazolium chloride). The insolubility of the product of these detection system posed an analytical challenge. To address this, another tetrazolium salt with a reportedly soluble product, XTT (2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide, was evaluated. While XTT yielded a soluble bright red product, the substrate was insoluble which thus effected the same analytical challenges. As an alternative to the tetrazolium family of dyes, the dual



colorimetric/fluorometric dye resazurin was evaluated. Oxidation of resazurin produces resourfin. Both substrate and product are soluble, and the color change can be quantified colorimetrically or fluorimetrically, increasing accuracy. Due to the sensitivity of resazurin, 0.05 mM of lactic acid can be quantified. Optimal results were obtained when using the dye in the same range as the substrate, *e.g.* 0.5 mM resazurin can quantify a range of lactic (and analogs) from 0.05 to 0.5, though the best linearity is at the lower end of this scale. Resourfin was stable over 28 hours, and had a linear fluorescent response.

[000460] In the presence of the LDH assay components, lyophilized enzyme gave background fluorescence/absorption. To address this problem the lysate was boiled for 10 minutes and then centrifuged. This resulted in a 90% decrease in background signal. Interestingly, both centrifugation alone (5 minutes @ 14.1 rcf) or boiling followed by centrifugation (5 minutes @ 100°C) reduced the fluorescence to background levels. In a high-throughput format such as 1536 well plates, spinning would be preferable to boiling, as boiling would increase evaporation (8 µl well size) and potentially volatilize the nitrile substrates. No background signal resulting from growth media (LB and TB and M9) or cell lytic solutions (B-PER, CellLytic and BugBuster) was noted.

#### Chiral AAO Assay

[000461] A spectroscopic system based on amino acid oxidase (AAO) was developed for the analysis of the chiral  $\alpha$ -amino acids which are generated by the nitrilase catalyzed hydrolysis of amino nitriles.

#### *Assay Development and Validation*

[000462] The initial assay validation utilized the 2,2'-azino-di-{3-ethylbenzothiazoline-6-sulfonic acid (ABTS) detection system as outlined above. However, since the color was not stable further investigations utilized the phenol amino antipyrine (PAAP) detection system which is analyzed at  $\lambda$  max 510nm. Enzymes with suitable activity were found for each enantiomer of 4-methyl-leucine, phenylalanine, (S)-2-amino-6-hydroxy hexanoic acid, and *tert*-leucine. The assay is not applicable to methylphenylglycine and does not work well with phenylglycine.

[000463] Standard curves were generated for phenylalanine from 0-15 mM. The curve is much more linear when the concentrations remained below 1 mM. The color remains stable for several days as long as it is kept in the dark. Three cell lysing solutions Bug Buster



(BB), Bacterial Protein Extracting Reagent (BPER), and Cell Lytic Reagent (CLR) were added to the standard curve and shown to have no affect on color development. The addition of cell lysate (cl) did not exhibit background color formation. Addition of the phenylacetaldehyde aminonitrile sulfate (PAS) starting material also showed no effect on color formation.

[000464] The AAO system exhibits greater linearity at up to 1 mM substrate. The concentration of the AAO enzymes and of the acid substrate were adjusted to try to move the intersection of the L-AAO and D-AAO curves closer to the middle of the graph. Premixing the PAAP, the HRP, and the AAO was demonstrated to be effective and caused no change in observed activity establishing that the assay components may be added to the assay in a cocktail format.

[000465] A high level of background was observed for the AAO assay of whole cells and this was attributed to the L-amino acids present in the TB and LB growth media. Washing and resuspension of the cells in M9 media eliminated background. For all future experiments cells were grown in M9 media with 0.2% glucose. The lysed cells gave only a slightly better response than unlysed cells. Therefore, cell lysis is not necessary. SEQ ID NOS:187, 188 demonstrated activity on HPA in primary screening based on HPLC analysis.

[000466] The use of a fluorescent detection system which would permits implementation of the assay in ultra high throughput fashion such as 1536 well or gigamatrix format was investigated. The fluorescent reagent most applicable to our system is Amplex Red from Molecular Probes which produces the highly fluorescent resorufin ( $\lambda_{\text{ex}}$  545 nm;  $\lambda_{\text{em}}$  590nm) Standard curves for phenylalanine and (S)-2-amino-6-hydroxy hexanoic acid were established (0-100  $\mu\text{M}$ ).

[000467] In preparation for assay automation, nitrilase expressing cells were added into microtiter plate containing M9 0.2% glucose, 0.25 mM IPTG media by fluorescence activated cell sorting (FACS). Three nitrilase expressing subclones, and the empty vector control were evaluated: SEQ ID NOS:101, 102, SEQ ID NOS:187, 188, SEQ ID NOS:29, 30 and the empty vector. The viability of the cells following cell sorting proved to be inconsistent. Thus colony picking is currently being evaluated as an alternative method to add cells into microtiter plates. The evaporative loss from an uncovered 1536-well microtiter plate is approximately 30% per day in the robot incubator (incubator conditions: 37°C at 85% relative humidity (RH)). Incubation in the 95% RH incubator reduced evaporative loss to 1% per day.



[000468] The ability of the three subclones to grow in the presence of up to 3.5 mM of nitrile was established using HPA nitrile. Growth rates were only slightly retarded (less than 30%). Subclones grown in the presence of HPA were shown to express a nitrilase that catalyzes the formation of hydroxy norleucine (HNL) as established using the Amplex Red detection system. Only *S* was evaluated as the enzymes are *S*-selective. The reaction plate was read at 10 minute intervals, with 40 minutes showing the best linearity. While cell growth is significantly inhibited above 5 mM of HPA when the cells were grown at pH 7, growth was inhibited above 0.1 mM HPA for cells grown at pH 8.

[000469] In order to verify the AAO results by HPLC, a reaction was performed using high concentrations of HPA, up to 40 mM (due to HPLC detection challenges for (S)-2-amino-6-hydroxy hexanoic acid) and lyophilized cell lysate SEQ ID NOS:187, 188.

Comparison AAO and HPLC data for HNL

| [HNL]<br>mM | %ee |      | %conversion |      |
|-------------|-----|------|-------------|------|
|             | AAO | HPLC | AAO         | HPLC |
| 40          | 89% | 100% | 17%         | 18%  |
| 30          | 89% | 97%  | 29%         | 36%  |
| 20          | 86% | 97%  | 21%         | 34%  |
| 10          | 78% | 98%  | 13%         | 35%  |

[000470] In order to determine if conducting the screen at a lower concentration introduces a bias in the results compared to the 20 mM substrate range that was used for HPLC based screens, an experiment was performed with SEQ ID NOS:187, 188 using three concentration ranges. Each experiment was done in triplicate in order to remove any nonsystematic error.

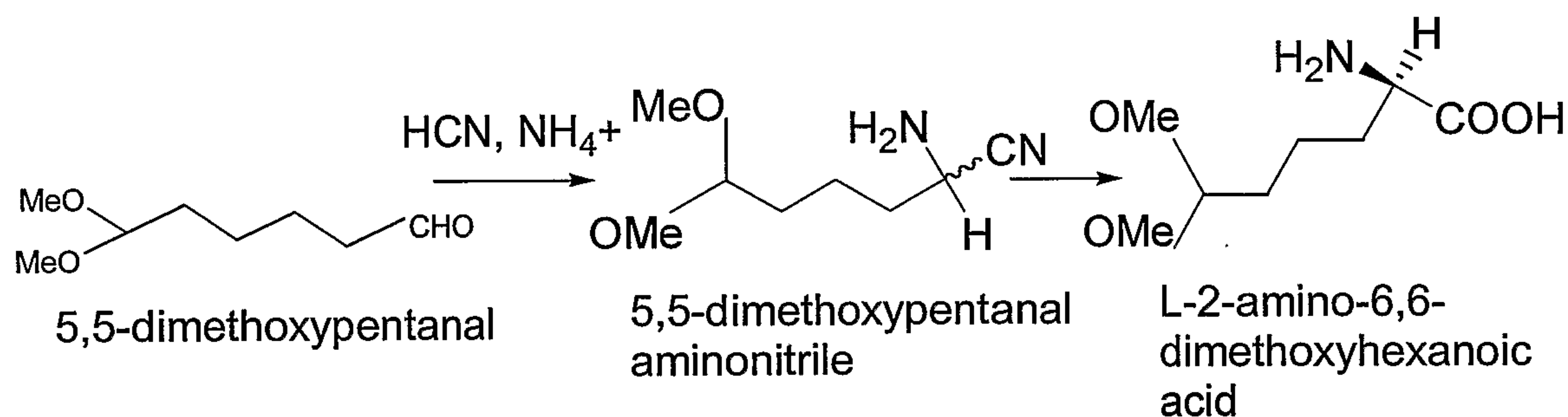
[000471] The AAO assay can be run on 384 or 1536 well format with cells sorted into an M9 0.2% glucose, 0.25 mM IPTG media. Cells can be grown in the presence of nitrile (in this case HPA), or the cells can be allowed to reach a certain density and the nitrile can then be added. Though cell lytic reagents do not interfere with the assay, when HPA was assayed, addition of the lytic reagents was found to be unnecessary. Either pre- or post- nitrile addition, the mother plate will have to be split into daughter plates, which are then assayed for the respective L- and D- enantiomer content. Incubation times with the AAO/Amplex Red reagents can be adjusted so that the D- and L- plate are read at separate times.



Example 16: Identification, Development and Production of Robust, Novel EnzymesTargeted for a Series of High-Value Enantioselective Bioprocesses

[000472] The invention provides for the development of nitrilases, through directed evolution, which provide significant technical and commercial advantages for the process manufacturing of the following chemical target:

L-2-amino-6,6-dimethoxyhexanoic acid



[000473]

Nitrilase enzymes were shown to hydrolyze 2-amino-6-hydroxy hexanenitrile to (S)-2-amino-6-hydroxy hexanoic acid, with selectivity towards the L-enantiomer. The panel of nitrilases was screened against the target, 5,5-dimethoxypentanal aminonitrile. The positive enzymes were characterized on this substrate. A primary screen is used to identify putative up-mutants, which is then confirmed using HPLC.

[000474]

GSSM<sup>TM</sup> and GeneReassembly<sup>TM</sup> are performed on selected nitrilases, in order to improve the enantioselectivity and activity of the enzymes for the production of L-2-amino-6,6-dimethoxyhexanoic acid. Nitrilases were identified for the enantioselective hydrolysis of 2-amino-6-hydroxy hexanenitrile to L-(S)-2-amino-6-hydroxy hexanoic acid. However, a slight structural difference is presented by the new target molecule, L-2-amino-6,6-dimethoxyhexanoic acid. In order to determine whether this difference affects the activity and enantioselectivity of the enzymes, the complete spectrum of nitrilases was screened against the new target.

[000475]

First, identification of the correct target gene for GSSM through more detailed characterization of the hit enzymes for the production of L-2-amino-6,6-dimethoxyhexanoic acid was carried out. This effort involves a more extensive investigation of the effects of pH



and temperature on activity and enantioselectivity and a more in-depth analysis of the stability of the enzyme to process conditions. Prior to initiation of the screening, the synthesis of a single enantiomer of an alkyl aminonitrile is done; the racemization of this nitrile is studied, in an effort to understand the relationship between this factor and

5 enantioselectivity of the enzymes.

An enzyme exhibiting the highest combination of activity and enantioselectivity for the production of L-2-amino-6,6-dimethoxyhexanoic acid is selected for GSSM. Following the mutation of the target enzyme, the resulting mutants are screened on 5,5-dimethoxypentanal aminonitrile, using high throughput screening technology. Following

10 confirmation of the up-mutants by HPLC analysis a decision point is reached, in order to evaluate the results of the GSSM on the target.

[000476] In parallel to GSSM<sup>TM</sup>, a GeneReassembly<sup>TM</sup> is performed on a combination of parent enzymes, at least one of which is selected for activity and enantioselectivity on L-2-amino-6,6-dimethoxyhexanoic acid. At least two other nitrilases are reassembled with the

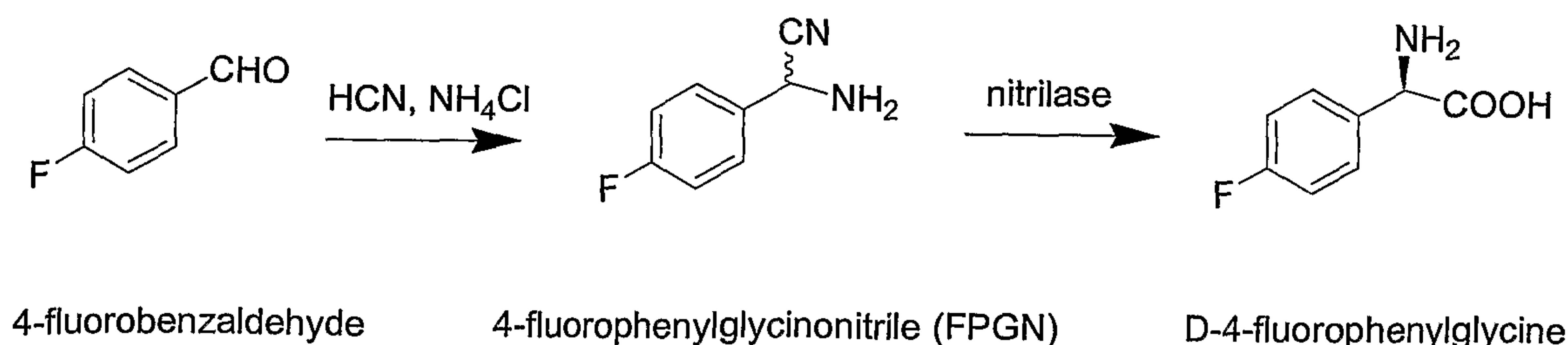
15 former enzyme(s); these enzymes are selected in order to provide diversity to the reassembled sequences.

[000477] The present invention provides for development of racemization conditions for the original substrate aminonitriles. In addition, the present invention provides for the identification of enzymes capable of the conversion of these aminonitriles to the target  $\alpha$ -

20 amino acids by dynamic kinetic resolution. The present invention also provides for screening and development of a nitrilase-catalyzed kinetic resolution process for (R)-2-amino-6,6-dimethoxy hexanoic acid (allysine) production. (S)-2-amino-6-hydroxy hexanoic acid will be used as a model substrate for development of the kinetic resolution.

The target  $\alpha$ -amino acid products are shown below:

25 (i) D-4-Fluorophenylglycine



4-fluorobenzaldehyde

4-fluorophenylglycinonitrile (FPGN)

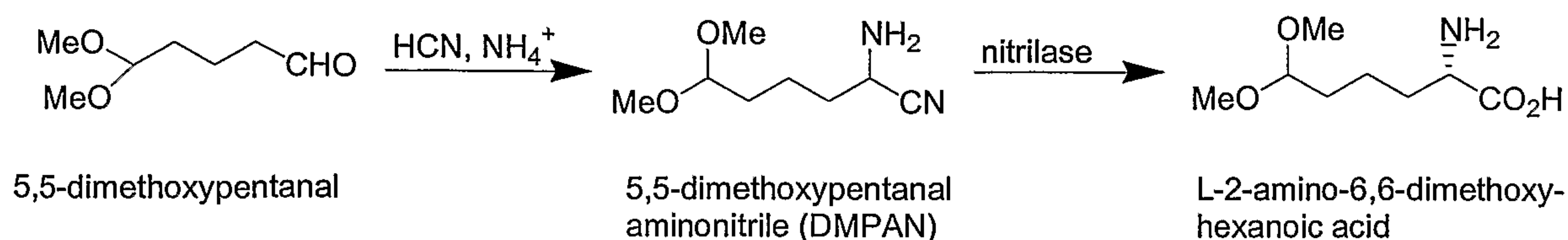
D-4-fluorophenylglycine

(ii) L-2-Amino-6,6-dimethoxyhexanoic acid (Allysine)



WO 03/097810

PCT/US03/15712



**[000478]** Conditions are developed for the racemization of the aminonitrile substrates for the nitrilase-catalyzed production of D-4-fluorophenylglycine and 2-amino-4,4-dimethyl pentanenitrile (allysine). Two model substrates, phenylglycinonitrile and pentanal

5 aminonitrile are used initially, and racemization is studied in the absence of the enzyme. Concurrently determination of the performance of one or more available nitrilases under a variety of possible racemization conditions is carried out. In addition, the nitrilases are screened against hydroxypentanal aminonitrile for the production of (S)-2-amino-6-hydroxy hexanoic acid, and the promising enzymes are optimized. Once racemization conditions are

10 established, the nitrilases are screened for activity. Further optimization for a kinetic resolution of the product is performed.

**[000479]** A number of enantioselective nitrilases were identified for the hydrolysis of  $\alpha$ -aminonitriles to  $\alpha$ -amino acids. While these enzymes were shown to have a preference for the required enantiomer of certain aminonitriles, a limiting factor in the further screening,

15 development and comparison of candidate nitrilases is the rate of racemization of the aminonitrile substrates under the reaction conditions.

#### Aromatic aminonitrile racemization

**[000480]** The first step is to establish conditions under which aromatic aminonitrile racemization occurs, using the model substrate, phenylglycinonitrile. Racemization strategies

20 include, but are not limited to the list below. Options are roughly prioritized according to their commercial applicability.

(1) Manipulation of the pH of the reaction. Since it has been shown that racemization is rapid at high pH, this approach requires the discovery and optimization of nitrilases which are active and selective at pH>10.

25 (2) Addition of known chemical racemizing agents, such as aldehydes, ketones, weak bases, resins, metal ions, Lewis acids etc., which can enhance racemization at lower pH.

(3) Synthesis of N-acylated aminonitrile derivatives, *e.g.* N-acetyl phenylglycinonitrile, which may be more easily racemized. In the case of N-acetyl phenylglycinonitrile, a selective D-acylase which removes the acetyl group would enhance the optical purity of the nitrilase product.

30 (4) Use of a biphasic system in which base-catalyzed racemization occurs in the hydrophobic organic phase and enzymatic hydrolysis in the aqueous phase.



(5) Use of a 2-enzyme system comprised of a nitrilase and an aminonitrile racemase. One amino acid racemase is commercially available at present, and will be tested for activity against phenyl- and fluorophenylglycinonitrile. Gene libraries will be searched for genes showing homology to known amino acid amide racemases, hydantoin racemases or any other racemases which can be identified.

5 [000481] Once conditions for this racemization have been established, they provide the basis for development of conditions for racemization of the target aromatic substrate, 4-fluorophenylglycinonitrile (FPGN). The FPGN is expected to be less stable than the model substrate; thus, it may racemize more quickly, but degradation reactions may be faster as well. The ability of sample enzyme(s) to tolerate and/or function well under them is  
10 evaluated. Final optimization of screening methods include the target substrates, sample nitrilases, and substrate racemization conditions.

[000482] Investigations carried out have shown that phenylglycinonitrile is easily racemized at pH 10.8. However, it does not appear that any of the existing enzymes can tolerate such harsh conditions of pH. Samples from highly alkaline environments are  
15 screened for the presence of nitrilases which are tolerant to such conditions. Once discovered, the enzymes are sequenced and subcloned, and the enzymes are produced as lyophilized cell lysates ready for screening.

#### Aliphatic aminonitrile racemization

[000483] A model aliphatic aminonitrile, pentanal aminonitrile, is synthesized in its  
20 racemic form. However optically enriched samples are prepared using one the following approaches: (i) preparative chiral HPLC; (ii) diastereomeric salt resolution; (iii) diastereomeric derivatization or column chromatography; (iv) synthesis from L-N-BOC norleucine. An HPLC assay is used for the detection of these compounds.

#### HPLC Assay

25 [000484] An HPLC assay for the detection of the (S)-2-amino-6-hydroxy hexanoic acid is used. An assay involving pre-column derivatization is used.

#### Screening/Characterization:

[000485] Nitrilases are screened against 2-amino-6-hydroxy hexanenitrile. For enzymes capable of performing well at greater than 25 mM substrate, scale up reactions are performed.  
30 The substrate/product tolerance and stability profiles of the other enzymes are investigated.

[000486] The nitrilases are screened, and hits are characterized, focusing on pH and temperature optimum, enantioselectivity and stability under the reaction conditions.



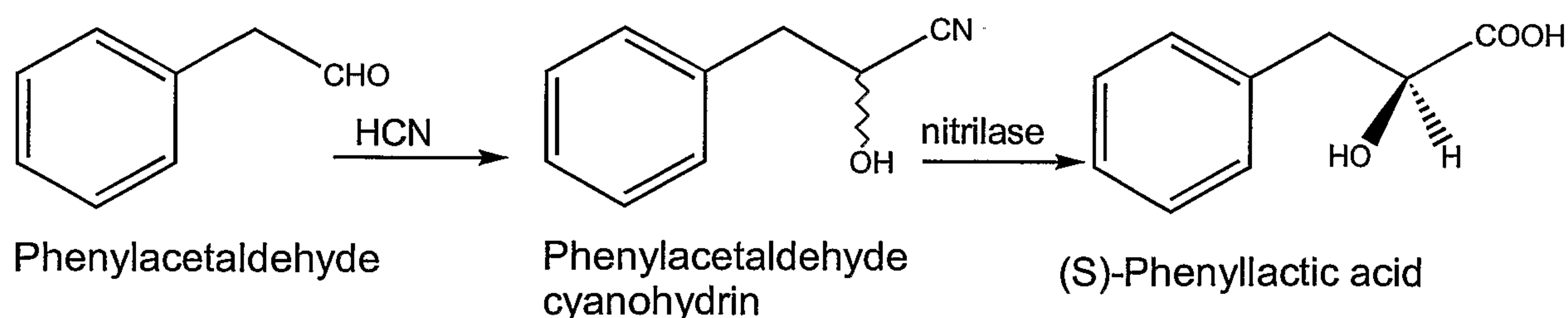
Enzyme Evolution

[000487] A target enzyme exhibiting the desired properties is selected for GSSM™. Following the mutation of the target enzyme, the resulting mutants are screened on the substrate using high throughput screening technology. Once the up-mutants have been confirmed by HPLC analysis, the individual mutations responsible for increased performance may be combined and evaluated for possible additive or synergistic effects.

[000488] In addition, a GeneReassembly™ will be performed on a combination of lead enzymes, which are selected for their desirable characteristics, including activity, enantioselectivity and stability in the reaction.

Example 17: Optimization of Nitrilases for the Enantioselective Production of (S)-Phenyllactic Acid

[000489] Nitrilases were identified for the enantioselective hydrolysis of 5 different nitrile substrates. These nitrilases were isolated and optimized for selected targets. The optimization involves process optimization and directed evolution. In particular, enzymes specific for the production of (S)-phenyllactic acid were characterized and optimized. This was aimed primarily at improving the activity of the enzymes, while maintaining a high enantioselectivity. An investigation into the effects of process conditions on the enzymes was also performed.



[000490] The development of high throughput assays for screening of mutants from potential directed evolution efforts was accomplished. Two achiral and two chiral colorimetric assays that are amenable to high throughput screening were developed and used for nitrilase directed evolution.

[000491] SEQ ID NOS:103, 104 was identified as a highly enantioselective nitrilase for the production of (S)-phenyllactic acid. Characterization of SEQ ID NOS:103, 104 shows the optimum reaction pH and temperature to be pH 8 and 37°C, respectively; the reaction starting material, phenylacetaldehyde, and the product, phenyllactic acid showed no effect on



the enzyme activity up to levels of 5 mM and 30 mM, respectively. The scaled-up enzymatic reaction with an enantiomeric excess (*ee*) of 95%.

Example 18: Directed Evolution of a nucleic acid encoding a nitrilase enzyme.

[000492] The *nitB* gene (GenBank Accession No. AX025996, from *Alcaligenes*  
5 *faecalis*) was subjected to Gene Site Saturated Mutagenesis™ or GSSM™ to generate a  
library of single amino acid substitution mutants covering the entire enzyme. The sequence  
of the "parental" *nitB* gene used in the directed evolution is SEQ ID NO: 103, 104. A *nitB*  
mutant library was generated from carrying out GSSM™. This *nitB* mutant library was then  
screened for clones with increased whole cell hydroxymethylthiobutryonitrile (HMTBN,  
10 which is a nitrilase substrate) activity. The product of the nitrilase reaction on that substrate  
is hydroxymethylthiobutyric acid (HMTBA).

[000493] Assays were run at 35°C with 100mM HMTBN and 100mM K<sub>3</sub>PO<sub>4</sub>, pH 7 to  
approximately 30-40% conversion. Two methods were used to quantitate HMTBN  
conversion, one being direct measurement of HMTBS produced by HPLC analysis and the  
15 other being indirect detection of residual HMTBN using the fluorescent cyanide assay, which  
has previously been described.

[000494] Putative *nitB* up mutants were subjected to a secondary assay to confirm the  
increased activity. In the secondary assay, up mutants and the wild type control were induced  
in expression medium in shake flasks. Shake flask cultures are then washed with 100mM  
20 K<sub>3</sub>PO<sub>4</sub>, pH7 and resuspended to the same optical density at 660nm. Kinetic assays were then  
performed with the normalized cell resuspensions under the same conditions used in the  
initial assays. Putative up mutants confirmed to have increased HMTBN activity were  
sequenced and tested for increased activity after transformation back into the same expression  
strain to ensure that increases in activity are not due to host mutations.

25 [000495] A confirmed *nitB* GSSM™ up-mutant is *nitB* G46P, which contains a glycine  
(GGT) to proline (CCG) substitution at amino acid 46. The whole cell HMTBN activity of  
this mutant is approximately 50% greater than that of wild type NitB at both 25°C and 35°C.  
Upon identification of the beneficial G46P mutation, GSSM™ was used again to generate a  
pool of double mutants using the *nitB* G46P template. These mutants all contain the G46P  
30 mutation and an additional single amino acid substitution at a random site. The double  
mutants were assayed for HMTBN activity greater than that of *nitB* G46P. Double, triple and  
quadruple mutants were created in order to speed up the mutation process and identify  
beneficial mutations more quickly. After the first few beneficial mutations were identified



and isolated, they were combined to generate double mutants, the best of which was DM18. DM18 was used as a template to generate triple mutants. The most active triple mutant was TM3 and that was used as a template to generate quadruple mutants. The most active quadruple mutant was QM2. The table summarizes these mutations.

| mutant | mutation 1        | mutation 2         | mutation 3        | mutation 4         |
|--------|-------------------|--------------------|-------------------|--------------------|
| DM18   | R (gcg) 29 C(tgt) | Y(tac) 207 M (atg) |                   |                    |
| TM3    | R (gcg) 29 C(tgt) | Y(tac) 207 M (atg) | L(ctt) 170 T(act) |                    |
| QM2    | R (gcg) 29 C(tgt) | Y(tac) 207 M (atg) | L(ctt) 170 T(act) | A(gcg) 197 N9(aat) |

5

[000496] The mutants were characterized first by studying their whole cell HMTBN activity. At 100 mM HMTBN, the HMTBS production rate of QM2 is 1.2 times greater than that of the parental gene. However, at 200 mM HMTBN, the rate of QM2 is 3.6 times that of the parental gene. The productivity of these mutants is increased considerably when the HMTBN concentration is raised from 100 mM to 300 mM. As to conversion rates, TM3 completely converted the substrate after 270 minutes and both DM18 and SM show greater than 75% conversion after this time. To further address the issue of HMTBN concentration effects on activity/productivity of NitB, several mutants were assayed at both 400 mM and 528 mM HMTBN. NitB is essentially inactive at these substrate concentrations, however the mutants retain significant activity at these concentrations. In particular, the activity at these high concentrations were essentially the same as their activity at 200 mM substrate. Therefore, the mutants can be used over a wide substrate concentration range and provide much more flexibility in utility than the NitB parental gene.

[000497] The mutants were shown to have higher expression levels than the parental gene and it also appeared tha the QM2 and TM3 mutants contained a greater proportion of soluble enzyme than the wild type as seen in SDS-PAGE analysis. As to stability, all of the enzymes showed essentially the same stability pattern at both 25°C and 35°C.

[000498] Finally, the mutants were subjected to codon-optimization. The approach was to optimize the codons and therefore increase the expression levels in the particular host cell. That would, in turn, increase the activity per cell of the enzyme. This resulted in increased whole cell activity in the codon-optimized mutants as compared to controls. The increase in activity was approximately double the activity. An *E. coli* expression system was used.

#### Example 19: Selected Examples of Compounds Produced From a Nitrilase-Catalyzed Reaction



[000499] The compounds listed in Figure 15 are selected compounds that can be produced from a nitrilase-catalyzed reaction using an enzyme and/or a method of the invention.

[000500] In addition, the following are potential products which can be made via the nitrilase Strecker format. More than 100 amino acids and many new drugs can be produced

5 from their respective aldehydes or ketones utilizing the nitrilase enzymes of the invention.

For example, large market drugs which can be synthesized using nitrilases of the invention include homophenylalanine, VASOTEC™, VASOTERIC™, TECZEM™, PRINIVIL™,

PRINZIDE™, ZESTRIL™, ZESTORETIC™, RAMACE™, TARKA™, MAVIK™,

TRANDOAPRIL™, TRANDOLAPRILAT™, ALTACE™, ODRIK™, UNIRETIC™,

10 LOTENSIN™, LOTREL™, CAPOTEN™, MONOPRIL™, TANATRIL™, ACECOL™,

LONGEST™, SPIRAPRIL™, QUINAPRIL™, and CILAZAPRIL™. Other chiral drugs

include DEMSER™ (alpha-methyl-L-Tyrosine), ALDOCHLOR™, LEVOTHROID™,

SYNTHROID™, CYTOMEL™, THYOLAR™, HYCODAN™, CUPRIMINE™,

DEPEN™, PRIMAXIN™, MIGRANOL™, D.H.E.-45, DIOVAN™, CEFOBID™, L-

15 DOPA, D-DOPA, D-alpha-methyl-DOPA, L-alpha-methyl-DOPA, L-gamma-

hydroxyglutamate, D-gamma-hydroxyglutamate, 3-(2-naphthyl)-L-alanine, D-homoserine,

and L-homoserine.

[000501] Furthermore, the nitrilase enzymes of the invention can be useful for synthesizing the following amino acids. Many of these amino acids have pharmaceutical applications. D-

20 phenylglycine, L-phenylglycine, D-hydroxyphenylglycine, L-hydroxyphenylglycine, L-

tertiary leucine, D-tertiary leucine, D-isoleucine, L-isoleucine, D-norleucine, L-norleucine,

D-norvaline, L-norvaline, D-2-thienylglycine, L-2-thienylglycine, L-2-aminobutyrate, D-2-

aminobutyrate, D-cycloleucine, L-cycloleucine, D-2-methylphenylglycine, L-2-

methylphenylglycine, L-thienylalanine, and D-thienylalanine.

25 [000502] The enzymes of the nitrilase enzymes of the invention can be useful for the

synthesis of the following natural amino acids: glycine, L-alanine, L-valine, L-leucine, L-

isoleucine, L-phenylalanine, L-tyrosine, L-tryptophan, L-cysteine, L-methionine, L-serine, D-

serine, L-threonine, L-lysine, L-arginine, L-histidine, L-aspartate, L-glutamate, L-asparagine,

L-glutamine, and L-proline. The following are examples of unnatural amino acids which can

30 be produced using the nitrilase enzymes of the invention. D-alanine, D-valine, D-leucine, D-

isoleucine, D-phenylalanine, D-tyrosine, D-tryptophan, D-cysteine, D-methionine, D-

threonine, D-lysine, D-arginine, D-histidine, D-aspartate, D-glutamate, D-asparagine, D-

glutamine, and D-proline.



[000503] Furthermore, nitrilase enzymes of the invention can be used in non-Strecker chemical reactions including the synthesis of more chiral drugs such as TAXOTERE™ as well as chiral drugs containing 3-hydroxy-glutaronitrile (a \$5.5B market); LIPITOR™, BAYCOL™, and LESCOT™. Chiral product targets that are not drugs include

5 PANTENOL™, L-phosphinothricin, D-phosphinothricin, D-fluorophenylalanine, and L-fluorophenylalanine. Finally, nitrilase can be used to produce unnatural amino acid compounds lacking a chiral center such as sarcosine, iminodiacetic acid, EDTA, alpha-aminobutyrate, and beta-alanine.

10 Figure 16 examples of substrates and products produced by the nitrilases of the invention and/or the methods of the invention. The chemical structures of the substrates and of the products are shown. The chemical reactions shown here are non-limiting examples of activities of the nitrilases of the invention.

Example 20: Exemplary Preparation Using a Polypeptide of a Variant of SEQ ID NO:210

[000504] The variant, nitrilase 1506-83-H7A, is SEQ ID NO:210 with the Ala at residue 15 190 replaced with His. At the codon level, the mutation that occurred was GCT to CAT. This variant exhibits improved enantioselectivity in the conversion of 3-hydroxyglutarylnitrile (HGN) to (R)-4-Cyano-3-hydroxybutyrate.

[000505] This variant has been demonstrated to perform this transformation in 100 mM pH 7 sodium phosphate buffer at room temperature. This mutant can perform in other buffer 20 systems and temperatures as well with the potential for providing additional altered properties. Exemplary properties include, but are not limited to, altered rates of the reaction, % ee, and stability. In particular, the altered properties can be a higher reaction rate, a higher % ee, and greater stability. Altered properties can be an increase or decrease of at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% more than 25 wildtype.

[000506] This variant was shown to perform the transformation by producing products in high enantiomeric excess of 10 mM to 3 M substrate (HGN). Higher or lower substrate concentrations are also possible. Enantiomeric excess greater than or equal to 95% have been achieved. However, enantiomeric excess can be at least 25%, 30%, 35%, 40%, 45%, 50%, 30 55%, 60%, 65%, 70%, 75%, 80%, 85%, or 90% more than wildtype.

[000507] Variants of the SEQ ID NOs: of the invention, can be cloned into expression vectors. For example, variants of nucleic acid sequence SEQ ID NO:195, 205, 207, 209, OR 237, and nucleotides that encode the variants of amino acid sequence SEQ ID NO:210 can



be cloned into exemplary vectors that include, but are not limited to, pSE420 (*E. coli* expression vector) and pMYC (*pseudomonas* expression vector).

Example 21: Preparation using variants of the invention:

[000508] Add 3-Hydroxyglutaronitrile (1 g, 9 mmol) drop-wise to a stirred solution of nitrilase cell lysate (normalized for 150 mg protein content) in 2.12 mL of 100 mM pH 7 sodium phosphate buffer at room temperature, ~22 °C. Stir this 3 M reaction by magnetic stir bar for 24 hours at room temperature. Monitor the progress of the reaction by TLC (Thin Layer Chromatography) and GC (Gas Chromatography). The reaction should be complete within 24 hours.

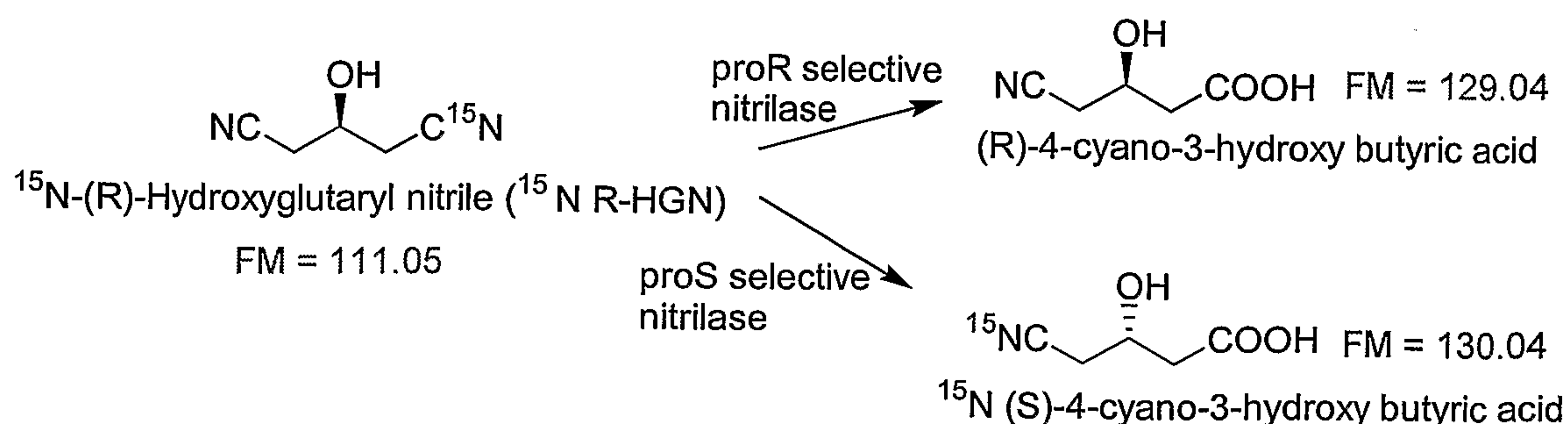
[000509] Other variants contemplated herein include, but are not limited to the following: N111S; A190H, S, Y or T; F191L, V, M, D, G, E, Y or T; M199E, orL; D222L; A55K, G, or Q; I60E, or any combination thereof.

Example 22: Screening Assay for Enantioselective Transformation

[000510] A new method to screen for enantioselective transformation, for example, of a prochiral substrate into a chiral one that affords the ability to monitor enantiomeric excess (% ee) of the resultant product is disclosed. This approach can also be applied to determine diastereomeric excess (% de).

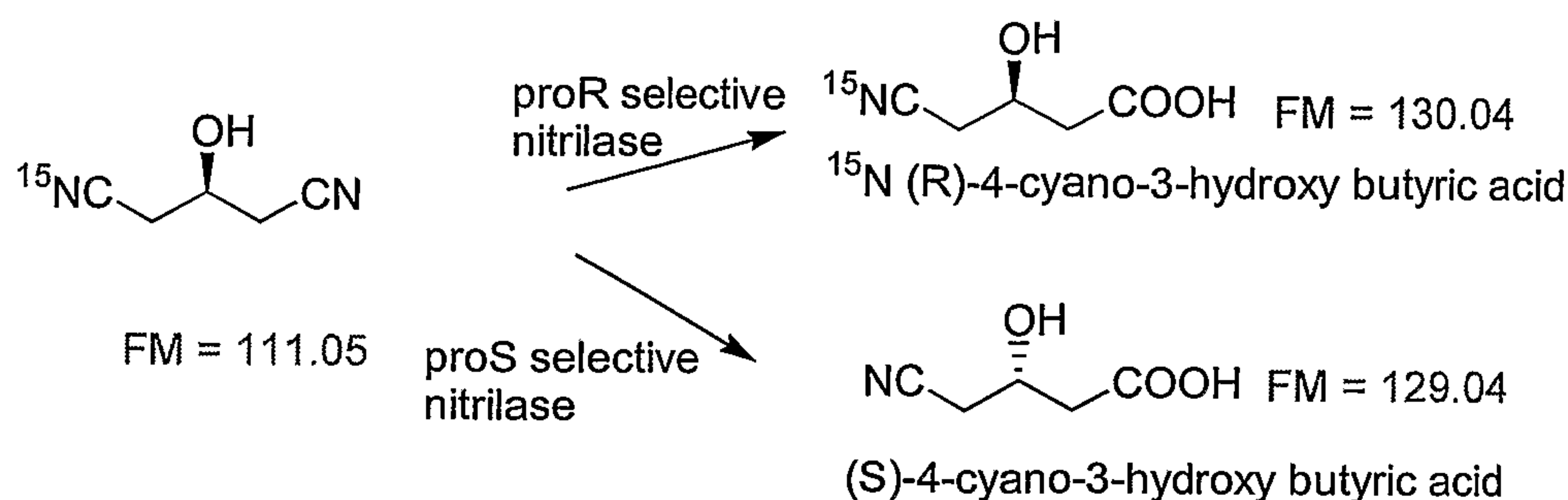
[000511] For example, by labeling one of the two prochiral or enantiotopic moieties in a molecule, for example by the use of a heavier or lighter isotope, the modification of one of the two moieties by a selective catalyst, for example, an enzyme, can be established by mass spectroscopy (MS).

[000512] By performing the exemplary nitrilase reaction on  $^{15}\text{N}$ -(R)-HGN (R) (as shown in Figure 17) or  $^{15}\text{N}$ -(S)-HGN, one can determine the enantioselectivity of the enzyme by analyzing the amount of each of the two possible labeled versus unlabeled acid products which can be formed.





[000513] The screening experiment may be performed in either direction. The screening experiment can be used for both the  $^{15}\text{N}$ -(R)- and (S)-HGN moieties. In fact, to ensure that the label does not effect any artifactual changes, at the onset, both should be investigated.

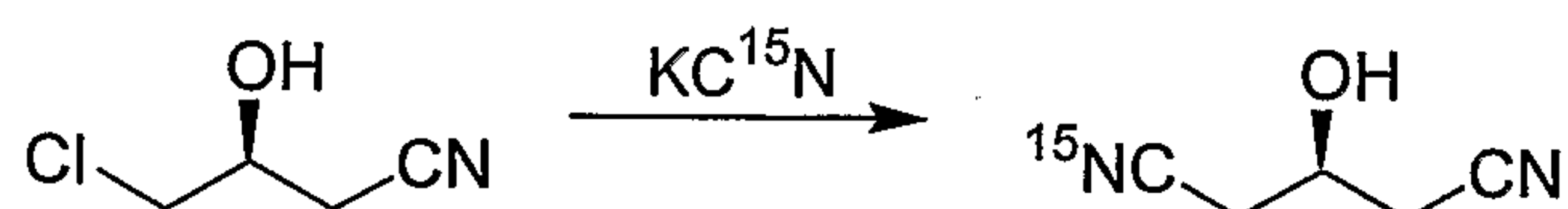


[000514] To equate the observed enantiomeric excess resulting from the nitrilase transformation, the following exemplary formula may be applied:

$$\% \text{ ee} = \{[130] - [129]\} / \{[130] + [129]\},$$
 where each concentration of the light acid (129) and the heavy acid (130) are determined by correlation of the peak area on the mass spectrometer to a standard curve or by direct comparison of the areas of each of the 129 and 130 mass peaks. The actual mass units used to determine the relative amounts of each of the two enantiomers (labeled and unlabelled) are dependent on how the mass spectrometer is tuned.

[000515] In some cases, the % ee observed by mass spectrometry may differ by a factor from that observed by an alternate analytical technique such as liquid chromatography due to background or contaminating peaks resulting from natural isotopic abundance. This does not, however, affect the final outcome of the screening process. Exemplary standard curves for quantification of heavy acid and light acid are shown in Figures 14 A and B.

[000516] The following reaction is a possible synthetic route to prepare, for example,  $^{15}\text{N}$  (R)-HGN using chemistry techniques known in the art with commercially available starting materials.



[000517] The amount of each of the two possible stereomeric outcomes can be established by the use of MS in either positive mode, negative mode and from analysis of either of the parental mass or of any fragmentation mass.



Example 23: Stability and activity of exemplary enzymes of the inventionEnzyme stability:

Wild-type enzyme (SEQ ID NOS:209 and 210) was compared to mutant A190H of SEQ ID NOS:209 and 210. In the experiment, each enzyme was incubated at 10 mg/ml in water for 1, 25, 50, 75 and 150 hours at 4°C and at 21°C, on two different substrates: adiponitrile and hydroxyglutaryl nitrile. Both enzymes, in all conditions, were found to retain activity for 150 hours. The wild-type enzyme showed greater activity on adiponitrile, while the mutated (A190H) enzyme showed greater activity on hydroxyglutaryl nitrile, as assessed by the Nitroprusside Bertholet assay (see, e.g., Fawcett, J. K. & Scott, J. (1960); J. Clin. Path.; Vol. 13, pg 156).

| GSSM <sup>TM</sup> Variant of<br>SEQ ID NOS:209 and<br>210 | 100 mM<br>hydroxyglutaryl<br>nitrile<br>ee% | 2.25 mM<br>hydroxyglutaryl<br>nitrile<br>ee% | Time to<br>completion<br>(hours) |
|------------------------------------------------------------|---------------------------------------------|----------------------------------------------|----------------------------------|
| A55G                                                       | 96.5 ± 0.4                                  | Not determined                               | >160                             |
| A55K                                                       | 94.7 ± 0.2                                  | Not determined                               | >160                             |
| I60E                                                       | 96.5 ± 0.5                                  | Not determined                               | >160                             |
| N111S                                                      | 95.8 ± 0.5                                  | 96.1 ± 0.9                                   | >160                             |
| A190T                                                      | 96.5 ± 0.2                                  | 96.6 ± 0.4                                   | 40                               |
| A190S                                                      | 96.8 ± 0.2                                  | 95.5 ± 0.7                                   | 40                               |
| A190H                                                      | 97.9 ± 0.1                                  | 98.1 ± 0.1                                   | 15                               |
| F191L                                                      | 97.9 ± 0.1                                  | Not determined                               | >160                             |
| F191T                                                      | 97.9 ± 0.1                                  | Not determined                               | >160                             |
| F191M                                                      | 97.9 ± 0.1                                  | Not determined                               | >160                             |
| F191V                                                      | 97.9 ± 0.1                                  | Not determined                               | >160                             |
| M199E                                                      | 97.9 ± 0.1                                  | Not determined                               | 160                              |
| M199L                                                      | 97.9 ± 0.1                                  | 95.4 ± 0.1                                   | >160                             |
| Wild type SEQ ID<br>NOS:209 and 210                        | 94.5 ± 0.1                                  | 87.8 ± 0.2                                   | 24                               |

GSSM mutants with enhanced enantioselectivity

100 mM reactions were performed with nitrilase expressed from E. coli in whole cell format and were complete with 36 hours. 2.25 M reactions were performed with nitrilase as lyophilized clarified cell lysate. All % ee data reported are the average of three measurements, with standard deviation of the mean. The time for reaction completion was approximated by TLC.

Specifically:Nitrilase Activity Assay, 100 mM HGN:



Putative nitrilase up-mutants were assayed in triplicate. Each transformant was grown in 5 mL LB (100 µg/mL ampicillin), at 37 °C, 220 rpm for 18 h. The overnight culture was diluted 2-fold and nitrilase expression induced at 37 °C, 220 rpm with 0.1 mM IPTG for 6 h. Cells were harvested by centrifugation, washed in 100 mM pH 7 sodium phosphate buffer and then re-suspended in 1 mL of 100 mM HGN in 100 mM pH 7 sodium phosphate buffer. Reactions were allowed to proceed for at least 36 h at 22 °C with gentle agitation. Reaction progress was monitored by TLC (1:1 EtOAc:Hexanes, R<sub>f</sub>=0.5, nitrile; R<sub>f</sub>=0.0, acid). Cells and other debris were removed by centrifugation and the treated with one volume methanol prior to lyophilization. The lyophilizate was re-suspended in methanol and treated with TMS-diazomethane (10 equivalents, 2 M solution in hexanes) until gas evolution ceased and yellow color persisted in order to prepare the methyl ester for GC analysis. Selected nitrilase variants producing (R)-(-)-3-hydroxy-4-cyanobutyric acid of 95% ee or greater were then evaluated for performance at 2.25 M HGN.

Nitrilase Activity Assay at 2.25 M 3-HGN:

3-HGN (0.2 g, 1.8 mmol, 3 M) was suspended in sodium phosphate buffer (0.6 mL, pH 7, 100 mM) at 22 °C. Cell lysate (6 mg, normalized for nitrilase content) was added to bring the concentration to 11 mg/ml enzyme and the reaction shaken (100 rpm, 22 °C). Reaction progress was monitored by TLC (1:1 ethyl acetate:hexanes, R<sub>f</sub> = 0.32, nitrile; R<sub>f</sub> = 0.0, acid). The reaction mixture was treated with one part methanol prior to lyophilization. The lyophilizate was re-suspended in methanol and treated with 10 equivalents of TMS-diazomethane (10 equivalents, 2 M solution in hexanes) to prepare the methyl ester and analyzed by GC.

Description of novel high throughput LC/MS method for screening high numbers of samples:

Ultra High-throughput Primary Chiral Activity Screen:

Distinct members of the GSSM library were arrayed into 384 well plates containing 40 µL of (Luria-Bertani) LB medium (100 µg/mL ampicillin) via an automated colony picker and then incubated at 37 °C, 85% humidity. Nitrilase expression was induced with 0.1 mM IPTG at 37 °C for 24 h. Each plate was replicated and 20% glycerol stocks prepared for archival at -80 °C. To each 384 well plate was added 10 mM 15N-(R)-1 substrate. The plates were incubated at 37 °C, 85% humidity for three days. Cells and other debris were removed by centrifugation and the supernatant was diluted 17,576-fold prior to MS analysis.



WO 03/097810

PCT/US03/15712

LC/MS ionspray was applied for high through-put analysis in the following manner. High-throughput screening was achieved by flow injecting samples from 384-well plates using a CTCPAL autosampler (Leap Technologies, Carrboro, N.C.). An isocratic mixture of 71% acetonitrile, 29% water, with 0.1% formic acid, provided by LC-10ADvp pumps (Shimadzu, Kyoto, Japan) at 2.2 mL/min through an LC-18 cartridge (Supelco, Bellefonte, PA) was used. Samples were applied to an API 4000 TurboIon spray triple-quadrupole mass spectrometer (Applied Biosystems, Foster City, CA). Ion spray and Multiple Reaction Monitoring (MRM) were performed for analytes in the negative ion mode, and each analysis took 60 seconds.

*E. coli* transformed with wild type enzyme (SEQ ID NOS:209 and 210) was used as a positive activity control and *E. coli* transformed with empty vector was used as the negative activity control. The % ee of the WT enzyme positive control determined by mass spectrometry using either 15N-(R)-1 or 15N-(S)-1 were the same, thus demonstrating the absence of a significant isotope effect.

| Temp (°C) | pH | Sodium phosphate buffer conc. (mM) | % ee | Std. Dev. |
|-----------|----|------------------------------------|------|-----------|
| 4         | 7  | 100                                | 98.7 | 0.1 %     |
| 19        | 7  | 100                                | 98.7 | 0.1 %     |
| 21        | 7  | 100                                | 98.6 | 0.1 %     |
| 37        | 7  | 100                                | 98.4 | 0.1 %     |
| 21        | 7  | 100                                | 98.6 | 0.1 %     |
| 21        | 6  | 100                                | 98.6 | 0.1 %     |
| 21        | 8  | 100                                | 98.6 | 0.1 %     |
| 21        | 7  | 100                                | 98.5 | 0.1 %     |
| 21        | 7  | 50                                 | 98.6 | 0.1 %     |
| 21        | 7  | 25                                 | 98.7 | 0.1 %     |

Effect of reaction parameters on SEQ ID NOS:209 and 210 with the A190H mutation.

Reactions were performed at 3 M HGN concentration with 150 mg/ml protein (~49 mg/ml enzyme). % ee was determined by GC analysis in triplicate runs.



## SEQUENCE LISTING

<110> DIVERSA CORPORATION

<120> Nitrilases, Nucleic Acids Encoding Them and Methods for Making and Using Them

<130> Docket No. 09010-120WO4

<140> Unknown

<141> 2003-05-15

<150> US 10/241,742

<151> 2002-09-09

<150> US 10/146,772

<151> 2002-05-15

<150> US 60/309,006

<151> 2001-07-30

<150> US 60/351,336

<151> 2002-01-22

<150> US 60/300,189

<151> 2001-06-21

<150> US 09/751,299

<151> 2000-12-28

<150> US 60/254,414

<151> 2000-12-07

<150> US 60/173,609

<151> 1999-12-29

<160> 386

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 939

<212> DNA

<213> Unknown

<220>

<223> Obtained from an environmental sample

<400> 1

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| atggaaaagt atattaaagt cgccgcaatt cagatagcta caaaaatagc agattcaccc  | 60  |
| gtgaatatag aaaattgcga acgtttggca ttatcggcgg tcaatgaggg tgcgcgttgg  | 120 |
| attgctttgc cggagttctt caatacgggc gttagttgga acaaaaaaat tgccttggct  | 180 |
| attcagacgc ctgacggcaa ggctgcgatg ttcttacgcg acttttctgc aagacatcat  | 240 |
| gtattgatag gaggctcatt tctgtgcagg ttgccggatg gcagtgtgcg caaccgctat  | 300 |
| atgtgttatg ccaacggcgc tctcgtgggc aaacatgaca aagacctacc cacgatgtgg  | 360 |
| gaaaatgctt tttatgaagg tggggattcc agcgatattg ggggtgctggg aacatttgaa | 420 |
| aatacgcgcg ttggtgcagc cgtctgttgg gagttcatgc ggacgatgac tgcccggcgt  | 480 |
| cttcgcaatc aggtggatgt catcatgggt ggttcctgct ggtggagcat accgaccaat  | 540 |
| ttccccggtt ttgtgcaaaa gctgtgggaa cctggaaata gccgcaacgc gcttgctgcc  | 600 |



```

atacaggata atgcgcgtct cattggcgtg ccggttggtc atgcgcgtca ttgcggtgaa 660
attgagtgtc cgatgccagg attgccgata ggttacaggg ggttctttga gggtaacgcg 720
gccattgtga atgcagaagg tcagggtgctt gcgcacgagg gtgctggcga gggcgaagga 780
attgtttgcg cggagatttt accggtagcc aaatcaaaca ggtcggaaat tcccaatcgt 840
tactggttgc gctgcagagg ctttctacct atttttgcct ggcatcagca acgttggttg 900
ggaaggcatt ggtatttgcg caatgtgcgc aggacttaa 939

```

&lt;210&gt; 2

&lt;211&gt; 312

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 2

```

Met Glu Lys Tyr Ile Lys Val Ala Ala Ile Gln Ile Ala Thr Lys Ile
 1          5          10          15
Ala Asp Ser Pro Val Asn Ile Glu Asn Cys Glu Arg Leu Ala Leu Ser
          20          25          30
Ala Val Asn Glu Gly Ala Arg Trp Ile Ala Leu Pro Glu Phe Phe Asn
          35          40          45
Thr Gly Val Ser Trp Asn Lys Lys Ile Ala Leu Ala Ile Gln Thr Pro
          50          55          60
Asp Gly Lys Ala Ala Met Phe Leu Arg Asp Phe Ser Ala Arg His His
          65          70          75          80
Val Leu Ile Gly Gly Ser Phe Leu Cys Arg Leu Pro Asp Gly Ser Val
          85          90          95
Arg Asn Arg Tyr Met Cys Tyr Ala Asn Gly Ala Leu Val Gly Lys His
          100          105          110
Asp Lys Asp Leu Pro Thr Met Trp Glu Asn Ala Phe Tyr Glu Gly Gly
          115          120          125
Asp Ser Ser Asp Ile Gly Val Leu Gly Thr Phe Glu Asn Thr Arg Val
          130          135          140
Gly Ala Ala Val Cys Trp Glu Phe Met Arg Thr Met Thr Ala Arg Arg
          145          150          155          160
Leu Arg Asn Gln Val Asp Val Ile Met Gly Gly Ser Cys Trp Trp Ser
          165          170          175
Ile Pro Thr Asn Phe Pro Gly Phe Val Gln Lys Leu Trp Glu Pro Gly
          180          185          190
Asn Ser Arg Asn Ala Leu Ala Ala Ile Gln Asp Asn Ala Arg Leu Ile
          195          200          205
Gly Val Pro Val Val His Ala Ala His Cys Gly Glu Ile Glu Cys Pro
          210          215          220
Met Pro Gly Leu Pro Ile Gly Tyr Arg Gly Phe Phe Glu Gly Asn Ala
          225          230          235          240
Ala Ile Val Asn Ala Glu Gly Gln Val Leu Ala His Arg Gly Ala Gly
          245          250          255
Glu Gly Glu Gly Ile Val Cys Ala Glu Ile Leu Pro Val Ala Lys Ser
          260          265          270
Asn Arg Ser Glu Ile Pro Asn Arg Tyr Trp Leu Arg Cys Arg Gly Phe
          275          280          285
Leu Pro Ile Phe Ala Trp His Gln Gln Arg Trp Leu Gly Arg His Trp
          290          295          300
Tyr Leu Arg Asn Val Arg Arg Thr
          305          310

```

&lt;210&gt; 3

&lt;211&gt; 981

&lt;212&gt; DNA



&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 3

```

atggggaatt cgttcaagat cgcggtggta caagcctgtc cggctcttct ggatcgtggc      60
gcgacagtcg ccaaggcatg ccgcctgata gcggaggcag ccgcggcggg cgccagcctg      120
gtggtctttc cggaggcggt tgtgcccgga taccactgtt gggctctggt cattccggca      180
gggcattcgc aaccactgcg ggagttatac gccgaactgg tggggaacgc cgtggcggtg      240
ccgggcgatg ccaccgatcg gctttgcgcg gcagccagag aagccggcgt ggtagtggcg      300
atcggcataca atgaagtgaag cagcgaagcc agcggcacga cgatttaca tacgctgctg      360
tacatcggag cggacggcgc gattctgggc aaacaccgca aagtaatgcc gacgggcgga      420
gagcgcctgg tctgggcgct tggcgatggg agcgacctgg aggtctacga cctgcctttc      480
ggccgattgg gtggcctggt gtgctgggag aactacatgc ccctggcccg gtacgcgatg      540
tcggcatggg gaaccgagat ctacgtggct ccaacttggg atcgcggaga accgtggctg      600
tccacaatgc ggcataatgc gaaagaaggg cgatgctacg tagtgggatg ctgcagttgc      660
atgaaaattg acgatgtacc cgaccggctg gcgttcaaag ggaagtatct gtcgacggcc      720
gagggctggc tcaaccccgg cgatagcgta atcgctcgatc cggacggcaa gctgatcgcg      780
ggcccggcaa gcgagcagga gacgattctg tatgccgatg ccgaccggtc taagatcacc      840
gggcccaggt ggcagttgga tgtggccggc cactacgcgc ggccggatat cttcgaactg      900
atcgtgcacc gcgaacctaa gcgatttttg acgatagctc cgcgacgaa ggaggagcgg      960
gagcctgggc cggaggcctg a                                     981

```

&lt;210&gt; 4

&lt;211&gt; 326

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 4

```

Met Gly Asn Ser Phe Lys Ile Ala Val Val Gln Ala Cys Pro Val Phe
 1          5          10          15
Leu Asp Arg Gly Ala Thr Val Ala Lys Ala Cys Arg Leu Ile Ala Glu
          20          25          30
Ala Ala Ala Ala Gly Ala Ser Leu Val Val Phe Pro Glu Ala Phe Val
          35          40          45
Pro Gly Tyr Pro Leu Trp Val Trp Phe Ile Pro Ala Gly His Ser Gln
          50          55          60
Pro Leu Arg Glu Leu Tyr Ala Glu Leu Val Gly Asn Ala Val Ala Val
65          70          75          80
Pro Gly Asp Ala Thr Asp Arg Leu Cys Ala Ala Ala Arg Glu Ala Gly
          85          90          95
Val Val Val Ala Ile Gly Ile Asn Glu Val Asn Ser Glu Ala Ser Gly
          100          105          110
Thr Thr Ile Tyr Asn Thr Leu Leu Tyr Ile Gly Ala Asp Gly Ala Ile
          115          120          125
Leu Gly Lys His Arg Lys Val Met Pro Thr Gly Gly Glu Arg Leu Val
          130          135          140
Trp Ala Leu Gly Asp Gly Ser Asp Leu Glu Val Tyr Asp Leu Pro Phe
145          150          155          160
Gly Arg Leu Gly Gly Leu Leu Cys Trp Glu Asn Tyr Met Pro Leu Ala
          165          170          175
Arg Tyr Ala Met Ser Ala Trp Gly Thr Glu Ile Tyr Val Ala Pro Thr
          180          185          190
Trp Asp Arg Gly Glu Pro Trp Leu Ser Thr Met Arg His Ile Ala Lys
          195          200          205

```



Glu Gly Arg Cys Tyr Val Val Gly Cys Cys Ser Cys Met Lys Ile Asp  
 210 215 220  
 Asp Val Pro Asp Arg Leu Ala Phe Lys Gly Lys Tyr Leu Ser Thr Ala  
 225 230 235 240  
 Glu Gly Trp Leu Asn Pro Gly Asp Ser Val Ile Val Asp Pro Asp Gly  
 245 250 255  
 Lys Leu Ile Ala Gly Pro Ala Ser Glu Gln Glu Thr Ile Leu Tyr Ala  
 260 265 270  
 Asp Ala Asp Arg Ser Lys Ile Thr Gly Pro Arg Trp Gln Leu Asp Val  
 275 280 285  
 Ala Gly His Tyr Ala Arg Pro Asp Ile Phe Glu Leu Ile Val His Arg  
 290 295 300  
 Glu Pro Lys Arg Phe Leu Thr Ile Ala Pro Arg Thr Lys Glu Glu Arg  
 305 310 315 320  
 Glu Pro Gly Pro Glu Ala  
 325

<210> 5  
 <211> 1005  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 5  
 atgggtacca agcaccgagg cttcaaggcc gcagtgggtcc aggccgcgcc ggaatggctc 60  
 gatctcgacc gcaccgtcga caagaccatc gcgctgatcg aggaggccgc cggcgccggc 120  
 gcgaagctca ttgcgttccc ggaaacctgg attcccggct atccgtggca catctgggtc 180  
 ggcacgcccgg cgtggggcgat cageccgggc ttcgtgcagc gctacttcga caattcactg 240  
 gcctacgaca gcccgcaggc ccagcgcacg gcggacgccg cgaagaagaa caagatcacc 300  
 gtgggtgctcg gcctgtcgga gcgcgagggt ggcagccttt atatctcgca gtggctgatt 360  
 gggccgggacg gcgagaccat tgccaagcgc cgcaaactgc gccccacca cgtcgagcgc 420  
 accgtgttcg gcgatggcga cggcagccac atcgcggtgc acgagcgtgc tgacatcggc 480  
 cgctcggcg cgctgtgctg ctgggagcac atccagccgc tgaccaaata cgccatgtat 540  
 gcccagaacg agcaggtgca cgtcgccgcc tggccgagct tctcgatgta cgagccgttc 600  
 gccacgcgc tcggctggga agtcaacaat gcggcgagca agatctacgc cgtcgaaggc 660  
 tcgtgttttcg tgctcggcgc atgocgggtg atctcgcagg cgatggtcga cgaaatgtgc 720  
 gacaccgagg acaagcgggc gctgggtccat gccggcggcg gccacgcggt gatcttcggg 780  
 ccggacggca gatcgctggc ggacaagatt ccggagaccc aggaaggcct gctctatgcc 840  
 gacatcgacc tcggcgcaat tggcgtggcc aagaacgcgg ccgatccggc ggggcactac 900  
 tcgcgcccgg acgtgacgcg gctcctgttc aacaacaagc cggcgcgccg ggctcgagtat 960  
 ttctcgctgc cggtcgacgc ggtcgagacg ccgcccgcagc cctga 1005

<210> 6  
 <211> 334  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 6  
 Met Gly Thr Lys His Pro Ala Phe Lys Ala Ala Val Val Gln Ala Ala  
 1 5 10 15  
 Pro Glu Trp Leu Asp Leu Asp Arg Thr Val Asp Lys Thr Ile Ala Leu  
 20 25 30  
 Ile Glu Glu Ala Ala Gly Ala Gly Ala Lys Leu Ile Ala Phe Pro Glu  
 35 40 45



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Thr | Trp | Ile | Pro | Gly | Tyr | Pro | Trp | His | Ile | Trp | Val | Gly | Thr | Pro | Ala |  |
| 50  |     |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |  |
| Trp | Ala | Ile | Ser | Arg | Gly | Phe | Val | Gln | Arg | Tyr | Phe | Asp | Asn | Ser | Leu |  |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |  |
| Ala | Tyr | Asp | Ser | Pro | Gln | Ala | Gln | Arg | Ile | Ala | Asp | Ala | Ala | Lys | Lys |  |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |  |
| Asn | Lys | Ile | Thr | Val | Val | Leu | Gly | Leu | Ser | Glu | Arg | Glu | Gly | Gly | Ser |  |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |
| Leu | Tyr | Ile | Ser | Gln | Trp | Leu | Ile | Gly | Pro | Asp | Gly | Glu | Thr | Ile | Ala |  |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |
| Lys | Arg | Arg | Lys | Leu | Arg | Pro | Thr | His | Val | Glu | Arg | Thr | Val | Phe | Gly |  |
| 130 |     |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |
| Asp | Gly | Asp | Gly | Ser | His | Ile | Ala | Val | His | Glu | Arg | Ala | Asp | Ile | Gly |  |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |  |
| Arg | Leu | Gly | Ala | Leu | Cys | Cys | Trp | Glu | His | Ile | Gln | Pro | Leu | Thr | Lys |  |
|     |     |     |     | 165 |     |     |     | 170 |     |     |     |     |     | 175 |     |  |
| Tyr | Ala | Met | Tyr | Ala | Gln | Asn | Glu | Gln | Val | His | Val | Ala | Ala | Trp | Pro |  |
|     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |  |
| Ser | Phe | Ser | Met | Tyr | Glu | Pro | Phe | Ala | His | Ala | Leu | Gly | Trp | Glu | Val |  |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |  |
| Asn | Asn | Ala | Ala | Ser | Lys | Ile | Tyr | Ala | Val | Glu | Gly | Ser | Cys | Phe | Val |  |
| 210 |     |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |
| Leu | Gly | Ala | Cys | Ala | Val | Ile | Ser | Gln | Ala | Met | Val | Asp | Glu | Met | Cys |  |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |
| Asp | Thr | Glu | Asp | Lys | Arg | Ala | Leu | Val | His | Ala | Gly | Gly | Gly | His | Ala |  |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |  |
| Val | Ile | Phe | Gly | Pro | Asp | Gly | Arg | Ser | Leu | Ala | Asp | Lys | Ile | Pro | Glu |  |
|     |     |     | 260 |     |     |     | 265 |     |     |     |     |     | 270 |     |     |  |
| Thr | Gln | Glu | Gly | Leu | Leu | Tyr | Ala | Asp | Ile | Asp | Leu | Gly | Ala | Ile | Gly |  |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |  |
| Val | Ala | Lys | Asn | Ala | Ala | Asp | Pro | Ala | Gly | His | Tyr | Ser | Arg | Pro | Asp |  |
| 290 |     |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |  |
| Val | Thr | Arg | Leu | Leu | Phe | Asn | Asn | Lys | Pro | Ala | Arg | Arg | Val | Glu | Tyr |  |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |
| Phe | Ser | Leu | Pro | Val | Asp | Ala | Val | Glu | Thr | Pro | Pro | Gln | Pro |     |     |  |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     |     |     |  |

<210> 7  
 <211> 999  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 7

|            |            |            |             |            |            |     |
|------------|------------|------------|-------------|------------|------------|-----|
| atgccgagtg | attatcatgc | tccattcaaa | gtagcagttg  | tccaggcaac | tcccgtcttt | 60  |
| ctcgatcgca | gcgcgacgat | tgagaaggca | tgtgagctaa  | ttgcctgtgc | tggacgtgag | 120 |
| ggcgcacgtc | tgatcgtgtt | tcctgaagcg | ttcattccca  | cctatcccga | ttgggtctgg | 180 |
| accattccac | ctggggagat | gcggctgctt | ggcgaactct  | acacagagtt | gcttgccaat | 240 |
| gcggtcacga | tccccagtaa | tgcaacggat | aggctctgcc  | aggctgcgaa | acgagctgct | 300 |
| gcgtatgtgg | tcatgggaat | gaacgaacgc | aatatcgagg  | cgagtggag  | gagtccttat | 360 |
| aacaccctgt | tatacatcga | tgctcagggc | cagatcatgg  | gcaaacaccg | caagttgata | 420 |
| cccacagccg | gtgagcggct | catatgggcg | caaggagatg  | ggagtacatt | ccaggtctac | 480 |
| gatactcctc | tgggcaaact | gggagggctc | atctgctggg  | aaaactacat | gcctctggct | 540 |
| cgctatgcga | tgtatgcctg | gggcacgcag | atcttatgtcg | ccccgacatg | ggatcgtggc | 600 |
| aacctctggc | tctctactct | gcggcatatc | gctaaggagg  | gaggcgtcta | tggtcttggt | 660 |
| tgtagtatgg | tcatgcgcaa | gaatgacatt | cccgatcact  | ttgctttcaa | agagcagttt | 720 |
| tatgctactg | tggacgaatg | gatcaacgtt | ggtgacagcg  | ccattgtcca | tcccgagggg | 780 |
| aactttcttg | cgggaccggt | gcgccacaaa | gaagagattc  | tctatgcaga | acttgatcca | 840 |



```

cgccaatcgt gcgggtccggg atggatgctc gatgtggctg ggcactatgc acgccctgat      900
gtgtttgaat tgattgtcca cacagagatg cgacccatga tgaagcaaga agaggtagga      960
ggagaaaata catctgaggg aggtgtacga ttcttgtaa      999

```

<210> 8  
 <211> 332  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 8

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Pro | Ser | Asp | Tyr | His | Ala | Pro | Phe | Lys | Val | Ala | Val | Val | Gln | Ala | 1   | 5   | 10  | 15  |
| Thr | Pro | Val | Phe | Leu | Asp | Arg | Ser | Ala | Thr | Ile | Glu | Lys | Ala | Cys | Glu | 20  | 25  | 30  |     |
| Leu | Ile | Ala | Cys | Ala | Gly | Arg | Glu | Gly | Ala | Arg | Leu | Ile | Val | Phe | Pro | 35  | 40  | 45  |     |
| Glu | Ala | Phe | Ile | Pro | Thr | Tyr | Pro | Asp | Trp | Val | Trp | Thr | Ile | Pro | Pro | 50  | 55  | 60  |     |
| Gly | Glu | Met | Arg | Leu | Leu | Gly | Glu | Leu | Tyr | Thr | Glu | Leu | Leu | Ala | Asn | 65  | 70  | 75  | 80  |
| Ala | Val | Thr | Ile | Pro | Ser | Asn | Ala | Thr | Asp | Arg | Leu | Cys | Gln | Ala | Ala | 85  | 90  | 95  |     |
| Lys | Arg | Ala | Ala | Ala | Tyr | Val | Val | Met | Gly | Met | Asn | Glu | Arg | Asn | Ile | 100 | 105 | 110 |     |
| Glu | Ala | Ser | Gly | Arg | Ser | Leu | Tyr | Asn | Thr | Leu | Leu | Tyr | Ile | Asp | Ala | 115 | 120 | 125 |     |
| Gln | Gly | Gln | Ile | Met | Gly | Lys | His | Arg | Lys | Leu | Ile | Pro | Thr | Ala | Gly | 130 | 135 | 140 |     |
| Glu | Arg | Leu | Ile | Trp | Ala | Gln | Gly | Asp | Gly | Ser | Thr | Phe | Gln | Val | Tyr | 145 | 150 | 155 | 160 |
| Asp | Thr | Pro | Leu | Gly | Lys | Leu | Gly | Gly | Leu | Ile | Cys | Trp | Glu | Asn | Tyr | 165 | 170 | 175 |     |
| Met | Pro | Leu | Ala | Arg | Tyr | Ala | Met | Tyr | Ala | Trp | Gly | Thr | Gln | Ile | Tyr | 180 | 185 | 190 |     |
| Val | Ala | Pro | Thr | Trp | Asp | Arg | Gly | Asn | Leu | Trp | Leu | Ser | Thr | Leu | Arg | 195 | 200 | 205 |     |
| His | Ile | Ala | Lys | Glu | Gly | Gly | Val | Tyr | Val | Leu | Gly | Cys | Ser | Met | Val | 210 | 215 | 220 |     |
| Met | Arg | Lys | Asn | Asp | Ile | Pro | Asp | His | Phe | Ala | Phe | Lys | Glu | Gln | Phe | 225 | 230 | 235 | 240 |
| Tyr | Ala | Thr | Val | Asp | Glu | Trp | Ile | Asn | Val | Gly | Asp | Ser | Ala | Ile | Val | 245 | 250 | 255 |     |
| His | Pro | Glu | Gly | Asn | Phe | Leu | Ala | Gly | Pro | Val | Arg | His | Lys | Glu | Glu | 260 | 265 | 270 |     |
| Ile | Leu | Tyr | Ala | Glu | Leu | Asp | Pro | Arg | Gln | Ser | Cys | Gly | Pro | Gly | Trp | 275 | 280 | 285 |     |
| Met | Leu | Asp | Val | Ala | Gly | His | Tyr | Ala | Arg | Pro | Asp | Val | Phe | Glu | Leu | 290 | 295 | 300 |     |
| Ile | Val | His | Thr | Glu | Met | Arg | Pro | Met | Met | Lys | Gln | Glu | Glu | Val | Gly | 305 | 310 | 315 | 320 |
| Gly | Glu | Asn | Thr | Ser | Glu | Gly | Gly | Val | Arg | Phe | Leu |     |     |     |     | 325 | 330 |     |     |

<210> 9  
 <211> 945  
 <212> DNA  
 <213> Unknown



&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 9

```

atggcggctc acaagatcgc ggtgggttcag gcgcccagcg ttctcctcga tcgcgagggc      60
tcggtcgcgc gcgcggtcac gctgctcgac gaggcggcgg cggccggcgc ccgcctggtc      120
gtgtttccgg aggcctacat ccccggttac ccggactgga tctggcgcct gcgcccctac      180
ccggacgtca agctggccgc cgagctgcac gaacggctgc tcgccaacgc ggtggatctc      240
tccaccgacg tgctggcgcc ggtgctggcg gcggcggcgc gtcacgggct caccgtggtc      300
atgtgcgtgc aggagcgcgga cgccggattc agccgcgcca cactttacaa caccgcgctg      360
gtcatcgacg ccgccggcaa gatcgcgaac cggcaccgca agctcatgcc caccaacccc      420
gagcgaatgg tgtgggggatt cggtgacgcc tcggggctgc gggtggtgag cagccccgtc      480
gggcgggtgg gcacgctcct gtgctgggag agctacatgc ccctggcgcg ctgcgcgctc      540
tacgccgagg gggtcgagat ctacgtgacc ccgacctggg actacggcga aggctggcgc      600
gccagcatgc agcacatcgc ccgcgagggg cgctgctggg tggtgaccgc ttgcatgtgc      660
gtgcaggcgc gcgacgtgcc ggccgacttc cccgggcgcg cccagctcta ccccgacgag      720
gaggagtggg tgaaccccgg cgattcgctg gtggtcgacc ccggcggcaa gatcgtggcc      780
ggtccgatgt cgcgcgagaa ggggatcttg tacgcggaga tcgatccgga tcgcgtggcg      840
ggggcgcacc gctcgttcga cgtcgtgggc cactactcgc gtcccagcgt gttccggctg      900
gaggtcgatc ggacaccggc ggcgccgggtg agcttcaaaa aatga                        945

```

&lt;210&gt; 10

&lt;211&gt; 314

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 10

```

Met Ala Ala His Lys Ile Ala Val Val Gln Ala Pro Ser Val Leu Leu
 1          5          10          15
Asp Arg Glu Gly Ser Val Ala Arg Ala Val Thr Leu Leu Asp Glu Ala
 20          25          30
Ala Ala Ala Gly Ala Arg Leu Val Val Phe Pro Glu Ala Tyr Ile Pro
 35          40          45
Gly Tyr Pro Asp Trp Ile Trp Arg Leu Arg Pro Tyr Pro Asp Val Lys
 50          55          60
Leu Ala Ala Glu Leu His Glu Arg Leu Leu Ala Asn Ala Val Asp Leu
 65          70          75          80
Ser Thr Asp Val Leu Ala Pro Val Leu Ala Ala Ala Arg His Gly
 85          90          95
Leu Thr Val Val Met Cys Val Gln Glu Arg Asp Ala Gly Phe Ser Arg
100          105          110
Ala Thr Leu Tyr Asn Thr Ala Leu Val Ile Asp Ala Ala Gly Lys Ile
115          120          125
Ala Asn Arg His Arg Lys Leu Met Pro Thr Asn Pro Glu Arg Met Val
130          135          140
Trp Gly Phe Gly Asp Ala Ser Gly Leu Arg Val Val Ser Thr Pro Val
145          150          155          160
Gly Arg Val Gly Thr Leu Leu Cys Trp Glu Ser Tyr Met Pro Leu Ala
165          170          175
Arg Cys Ala Leu Tyr Ala Glu Gly Val Glu Ile Tyr Val Thr Pro Thr
180          185          190
Trp Asp Tyr Gly Glu Gly Trp Arg Ala Ser Met Gln His Ile Ala Arg
195          200          205
Glu Gly Arg Cys Trp Val Val Thr Ala Cys Met Cys Val Gln Ala Arg
210          215          220

```



Asp Val Pro Ala Asp Phe Pro Gly Arg Ala Gln Leu Tyr Pro Asp Glu  
 225 230 235 240  
 Glu Glu Trp Leu Asn Pro Gly Asp Ser Leu Val Val Asp Pro Gly Gly  
 245 250 255  
 Lys Ile Val Ala Gly Pro Met Ser Arg Glu Lys Gly Ile Leu Tyr Ala  
 260 265 270  
 Glu Ile Asp Pro Asp Arg Val Ala Gly Ala His Arg Ser Phe Asp Val  
 275 280 285  
 Val Gly His Tyr Ser Arg Pro Asp Val Phe Arg Leu Glu Val Asp Arg  
 290 295 300  
 Thr Pro Ala Ala Pro Val Ser Phe Lys Lys  
 305 310

<210> 11  
 <211> 966  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 11  
 atgactggat cttatcctaa agacacactg atcgttgggc tagctcaa at cgctcctgtc 60  
 tggctggatc gggcggggac actgtcaaag atactggctc aagtccatgc ggcaaatcaa 120  
 gcgggttgct atctcgtagc atttggcgaa gggctgcttc ctggatatcc gttttggatt 180  
 gagcgaacaa atggcgcgct gttcaactcg actgtacaaa aggaaatcca cgcgcattat 240  
 atggatcagg cgggtgcagat cgaagccggt catctcgatc cgctttgtgc aacagccaag 300  
 aaatttgaa tcaccgttgt actcggatgc atcgaacgcc cactcgatcg gggcgggtcac 360  
 agcttgatg caagtctggt atatattgat tccgagggca gcattcaatc cgtgcatcgc 420  
 aaactaatgc caacctacga agaacgactt acctggtcgt caggcgatgg gcacgggtta 480  
 cgagtgcata ccttaggtgc gtttacggtg ggtggtctca actggtggga aaattggatg 540  
 cccttggcgc gcgcagcgat gtatggtcag ggtgaagatt tacatgttgc gatctggcca 600  
 ggcggttctc atctcacgca ggatattacc cgctttattg cgctcgaatc acgttcgtac 660  
 gtattatctg tctccggtct gatgcgcgca accgattttc caaaagatac tccccatctt 720  
 gcctccatcc tagctaaagg tgaagagatt cttgcgaatg gtggttcttg tattgcaggt 780  
 cctgacggca agtgggtcgt tgggcctctt gtaggagaag agaagttaat tgctcgcaacc 840  
 attgatcact gccgcgtgcg cgaagaacgt cagaatttcg atccttccgg gcattacagc 900  
 cggcccgatg tactgcaatt aaaaatcaac agggaaacgcc agagcacaat ttcatttagc 966  
 gagtag

<210> 12  
 <211> 321  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 12  
 Met Thr Gly Ser Tyr Pro Lys Asp Thr Leu Ile Val Gly Leu Ala Gln  
 1 5 10 15  
 Ile Ala Pro Val Trp Leu Asp Arg Ala Gly Thr Leu Ser Lys Ile Leu  
 20 25 30  
 Ala Gln Val His Ala Ala Asn Gln Ala Gly Cys His Leu Val Ala Phe  
 35 40 45  
 Gly Glu Gly Leu Leu Pro Gly Tyr Pro Phe Trp Ile Glu Arg Thr Asn  
 50 55 60  
 Gly Ala Leu Phe Asn Ser Thr Val Gln Lys Glu Ile His Ala His Tyr  
 65 70 75 80



Met Asp Gln Ala Val Gln Ile Glu Ala Gly His Leu Asp Pro Leu Cys  
                                     85                                    90                                    95  
 Ala Thr Ala Lys Lys Phe Gly Ile Thr Val Val Leu Gly Cys Ile Glu  
                                     100                                    105                                    110  
 Arg Pro Leu Asp Arg Gly Gly His Ser Leu Tyr Ala Ser Leu Val Tyr  
                                     115                                    120                                    125  
 Ile Asp Ser Glu Gly Ser Ile Gln Ser Val His Arg Lys Leu Met Pro  
                                     130                                    135                                    140  
 Thr Tyr Glu Glu Arg Leu Thr Trp Ser Ser Gly Asp Gly His Gly Leu  
 145                                    150                                    155                                    160  
 Arg Val His Thr Leu Gly Ala Phe Thr Val Gly Gly Leu Asn Cys Trp  
                                     165                                    170                                    175  
 Glu Asn Trp Met Pro Leu Ala Arg Ala Ala Met Tyr Gly Gln Gly Glu  
                                     180                                    185                                    190  
 Asp Leu His Val Ala Ile Trp Pro Gly Gly Ser His Leu Thr Gln Asp  
                                     195                                    200                                    205  
 Ile Thr Arg Phe Ile Ala Leu Glu Ser Arg Ser Tyr Val Leu Ser Val  
                                     210                                    215                                    220  
 Ser Gly Leu Met Arg Ala Thr Asp Phe Pro Lys Asp Thr Pro His Leu  
 225                                    230                                    235                                    240  
 Ala Ser Ile Leu Ala Lys Gly Glu Glu Ile Leu Ala Asn Gly Gly Ser  
                                     245                                    250                                    255  
 Cys Ile Ala Gly Pro Asp Gly Lys Trp Val Val Gly Pro Leu Val Gly  
                                     260                                    265                                    270  
 Glu Glu Lys Leu Ile Val Ala Thr Ile Asp His Cys Arg Val Arg Glu  
                                     275                                    280                                    285  
 Glu Arg Gln Asn Phe Asp Pro Ser Gly His Tyr Ser Arg Pro Asp Val  
                                     290                                    295                                    300  
 Leu Gln Leu Lys Ile Asn Arg Glu Arg Gln Ser Thr Ile Ser Phe Ser  
 305                                    310                                    315                                    320  
 Glu

<210> 13  
 <211> 1014  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 13  
 atgggcattc aacatccgaa atataagggtt gcggtggtgc aggcggcgcc ggcctggctc 60  
 gatctcgatg cgtegatcgc caaatcgatc gcgttgatcg aggaggcggc tgccaatggc 120  
 gccaagctga tcgccttccc ggaggcggtt atccctggct atccctggta tatctggctg 180  
 gactcgccgg cctggggcgat cggccgcggt tttgtgcagc gctatttcga caactcgctg 240  
 gcctatgaca gcccgcaggg cgagaagctg cggctggcgg tgaagaaggc cggcctcacc 300  
 gccgtgatcg gcctctccga gcgcgagggc ggcagccttt atctcgcgca atggctgatc 360  
 gggcccgatg gcgagaccat cgcaaagcgc cgcaagctgc ggccgaccca tgccgagcgc 420  
 accgtctatg gcgaaggcga tggcagcgat ctgcgcggtg atgaccgccc cggcatcggc 480  
 cggctcggcg cgctgtgctg ctgggagcat ctgcagccgc tgtcgaaata cgcgatgtat 540  
 gcccagaacg agcaggttca tgtcgcggcc tggccgagct tctcgctcta cgaccggtt 600  
 gcgcccggcg tcggctggga ggtcaacaat gcggcctcac gcgtctatgc ggtggaaggc 660  
 tcgtgcttcg tgctggcgcc ctgcgcgacg gtgtcgaagg cgatgatcga cgagctctgc 720  
 gaccgcgacg acaagcacgg gctgctgcat gtcggcgggg gacacgcccg gatctatggg 780  
 ccggacggct cttegatattg ggagaaattg ccgccggagc aggagggcct gctctatgcc 840  
 gacatcgatc tcggcgccat cgggattgcc aagaacgccg ccgatccggc cggacattac 900  
 tcgcggcccc acgtgacgcg gctgttgctc aacaagaagc cgtcgaagcg tgtcgagcat 960  
 ttttcgctgc cggctcgacaa tgtcgagccg gagatcgacg ccgccgccag ctga 1014

<210> 14



WO 03/097810

PCT/US03/15712

<211> 337  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 14

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Gly | Ile | Gln | His | Pro | Lys | Tyr | Lys | Val | Ala | Val | Val | Gln | Ala | Ala | 1   | 5   | 10  | 15  |
| Pro | Ala | Trp | Leu | Asp | Leu | Asp | Ala | Ser | Ile | Ala | Lys | Ser | Ile | Ala | Leu | 20  | 25  | 30  |     |
| Ile | Glu | Glu | Ala | Ala | Ala | Asn | Gly | Ala | Lys | Leu | Ile | Ala | Phe | Pro | Glu | 35  | 40  | 45  |     |
| Ala | Phe | Ile | Pro | Gly | Tyr | Pro | Trp | Tyr | Ile | Trp | Leu | Asp | Ser | Pro | Ala | 50  | 55  | 60  |     |
| Trp | Ala | Ile | Gly | Arg | Gly | Phe | Val | Gln | Arg | Tyr | Phe | Asp | Asn | Ser | Leu | 65  | 70  | 75  | 80  |
| Ala | Tyr | Asp | Ser | Pro | Gln | Ala | Glu | Lys | Leu | Arg | Leu | Ala | Val | Lys | Lys | 85  | 90  | 95  |     |
| Ala | Gly | Leu | Thr | Ala | Val | Ile | Gly | Leu | Ser | Glu | Arg | Glu | Gly | Gly | Ser | 100 | 105 | 110 |     |
| Leu | Tyr | Leu | Ala | Gln | Trp | Leu | Ile | Gly | Pro | Asp | Gly | Glu | Thr | Ile | Ala | 115 | 120 | 125 |     |
| Lys | Arg | Arg | Lys | Leu | Arg | Pro | Thr | His | Ala | Glu | Arg | Thr | Val | Tyr | Gly | 130 | 135 | 140 |     |
| Glu | Gly | Asp | Gly | Ser | Asp | Leu | Ala | Val | His | Asp | Arg | Pro | Gly | Ile | Gly | 145 | 150 | 155 | 160 |
| Arg | Leu | Gly | Ala | Leu | Cys | Cys | Trp | Glu | His | Leu | Gln | Pro | Leu | Ser | Lys | 165 | 170 | 175 |     |
| Tyr | Ala | Met | Tyr | Ala | Gln | Asn | Glu | Gln | Val | His | Val | Ala | Ala | Trp | Pro | 180 | 185 | 190 |     |
| Ser | Phe | Ser | Leu | Tyr | Asp | Pro | Phe | Ala | Pro | Ala | Leu | Gly | Trp | Glu | Val | 195 | 200 | 205 |     |
| Asn | Asn | Ala | Ala | Ser | Arg | Val | Tyr | Ala | Val | Glu | Gly | Ser | Cys | Phe | Val | 210 | 215 | 220 |     |
| Leu | Ala | Pro | Cys | Ala | Thr | Val | Ser | Lys | Ala | Met | Ile | Asp | Glu | Leu | Cys | 225 | 230 | 235 | 240 |
| Asp | Arg | Asp | Asp | Lys | His | Gly | Leu | Leu | His | Val | Gly | Gly | Gly | His | Ala | 245 | 250 | 255 |     |
| Ala | Ile | Tyr | Gly | Pro | Asp | Gly | Ser | Ser | Ile | Ala | Glu | Lys | Leu | Pro | Pro | 260 | 265 | 270 |     |
| Glu | Gln | Glu | Gly | Leu | Leu | Tyr | Ala | Asp | Ile | Asp | Leu | Gly | Ala | Ile | Gly | 275 | 280 | 285 |     |
| Ile | Ala | Lys | Asn | Ala | Ala | Asp | Pro | Ala | Gly | His | Tyr | Ser | Arg | Pro | Asp | 290 | 295 | 300 |     |
| Val | Thr | Arg | Leu | Leu | Leu | Asn | Lys | Lys | Pro | Ser | Lys | Arg | Val | Glu | His | 305 | 310 | 315 | 320 |
| Phe | Ser | Leu | Pro | Val | Asp | Asn | Val | Glu | Pro | Glu | Ile | Asp | Ala | Ala | Ala | 325 | 330 | 335 |     |

Ser

<210> 15  
 <211> 1047  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample



<400> 15  
 atgccaacat caaaacaatt tagagtcgct gcagttcaag ccgccccggt atttcttgac 60  
 ctggagggcg caataagcaa aggcattctc ctcattgagg aggccgcttc caatggagcc 120  
 aagctcattg ccttcccgga aacgtggatt cccggctacc cctgggtggat ctggctggac 180  
 tcacccgctt ggggcatgcg ctttgtccag cgctattttg acaactcgct catgctgggt 240  
 agtgagcaag ccaagcgcat gaaccaggct gccgccaata acaagattta cgtgggtgatg 300  
 ggttatagcg aacgcagtgg cggcagcctc tacatgggccc aatccattat caacgacaag 360  
 ggtgaaacga tttttacccg ccgcaaactc aagccaactc atgtcgagcg taccgtgttt 420  
 ggggagggag acggcagcca tctttgogta atggataccg agattggccg cgtcggcgcg 480  
 atgtgctgtt gggaacattt gcagccgctc agcaaataatg caatgtattc tcaggatgaa 540  
 caaattcaca ttgcctcctg gccgagcttt tcgttatatc ggggggagcgc ctatgcactc 600  
 ggccccgaac tgaacaacgc cgccagccaa atgtatgcag ccgaaggcca gtgctttgtc 660  
 cttgcccctt gcgccaccgt ctcaaaggag atgatcgaaa tgctgataga tgatcccagg 720  
 aaagagccgc ttctgctgga aggtggcggg ttcacatga tttacggccc cgatgggcca 780  
 cctctggcta aaccgttgcc tgaaaacgag gaagggtgc tatatgccga tattgacctg 840  
 ggcatgattt caatggccaa ggctgccgcc gaccggcgag gtcactacgc acgcccggat 900  
 gtcactcgcc tactattcaa ttccgcgccc gccaatcgcg tcgagtatat caaccagcg 960  
 tcaggcccaa ccgaatcctt aaaagatatg ggaaagatgc aatggaggc cgaacagcaa 1020  
 aaggcggccc tgcgagagat gatctaa 1047

<210> 16  
 <211> 348  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 16  
 Met Pro Thr Ser Lys Gln Phe Arg Val Ala Ala Val Gln Ala Ala Pro  
 1 5 10 15  
 Val Phe Leu Asp Leu Glu Gly Ala Ile Ser Lys Gly Ile Ser Leu Ile  
 20 25 30  
 Glu Glu Ala Ala Ser Asn Gly Ala Lys Leu Ile Ala Phe Pro Glu Thr  
 35 40 45  
 Trp Ile Pro Gly Tyr Pro Trp Trp Ile Trp Leu Asp Ser Pro Ala Trp  
 50 55 60  
 Gly Met Arg Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu Met Leu Gly  
 65 70 75 80  
 Ser Glu Gln Ala Lys Arg Met Asn Gln Ala Ala Ala Asn Asn Lys Ile  
 85 90 95  
 Tyr Val Val Met Gly Tyr Ser Glu Arg Ser Gly Gly Ser Leu Tyr Met  
 100 105 110  
 Gly Gln Ser Ile Ile Asn Asp Lys Gly Glu Thr Ile Phe Thr Arg Arg  
 115 120 125  
 Lys Leu Lys Pro Thr His Val Glu Arg Thr Val Phe Gly Glu Gly Asp  
 130 135 140  
 Gly Ser His Leu Cys Val Met Asp Thr Glu Ile Gly Arg Val Gly Ala  
 145 150 155 160  
 Met Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys Tyr Ala Met Tyr  
 165 170 175  
 Ser Gln Asp Glu Gln Ile His Ile Ala Ser Trp Pro Ser Phe Ser Leu  
 180 185 190  
 Tyr Arg Gly Ala Ala Tyr Ala Leu Gly Pro Glu Leu Asn Asn Ala Ala  
 195 200 205  
 Ser Gln Met Tyr Ala Ala Glu Gly Gln Cys Phe Val Leu Ala Pro Cys  
 210 215 220  
 Ala Thr Val Ser Lys Glu Met Ile Glu Met Leu Ile Asp Asp Pro Arg  
 225 230 235 240  
 Lys Glu Pro Leu Leu Leu Glu Gly Gly Gly Phe Thr Met Ile Tyr Gly  
 245 250 255



Pro Asp Gly Arg Pro Leu Ala Lys Pro Leu Pro Glu Asn Glu Glu Gly  
 260 265 270  
 Leu Leu Tyr Ala Asp Ile Asp Leu Gly Met Ile Ser Met Ala Lys Ala  
 275 280 285  
 Ala Ala Asp Pro Ala Gly His Tyr Ala Arg Pro Asp Val Thr Arg Leu  
 290 295 300  
 Leu Phe Asn Ser Ala Pro Ala Asn Arg Val Glu Tyr Ile Asn Pro Ala  
 305 310 315 320  
 Ser Gly Pro Thr Glu Ser Leu Lys Asp Met Gly Lys Met Gln Met Glu  
 325 330 335  
 Ala Glu Gln Gln Lys Ala Ala Leu Arg Glu Met Ile  
 340 345

<210> 17  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 17  
 atgagagttg ttaaagccgc agctgtccaa ctgagtcccg tcctctatag ccgcgagggg 60  
 acggtcgaga aggtcgtgcg gaagatccat gaacttgccg aagagggagt cgagttcgtc 120  
 acctttcctg agaccgtggt gccttattac ccgtactttt cgttcgttca gacgcccttg 180  
 cagcaaattc tcggaacaga gtatctgagg ctgctcgacc aggcagtcac cgtgccatcc 240  
 gccgccaccg acgcgatcgg cgaggctgcc aggttcgctg gagttgttgt ctcgatcggc 300  
 gtcaacgagc gagacggggg aactctgtac aacactcagc ttctcttcga tgccgacgga 360  
 agcttaattc agcggcgccg caagatcacg cccacccatt acgagcgcac gatctggggc 420  
 cagggtgacg gctcaggtct gcgggcccgt gatagcaagg ccggccgcat tggtcagctg 480  
 gcatgctggg agcacaacaa tccactggcg cgctacgcgc tgatagccga cggcgagcag 540  
 atccattcgg ccatgtatcc gggctccatg ttcggcgact cgtttgccaa aaagaccgaa 600  
 atcaatatcc ggcagcatgc gctggagtct gcgtgcttcg tcgtgaacgc aacggcctgg 660  
 ctggacggcg atcaacaggc gcaaatcatg aaggacaccg gctgcagcat cggcccgatc 720  
 tccggcgggt gcttcaccac tatcgtggcg ccggacgggt ccctgatcgg cgagcccctc 780  
 cgctcgggtg agggcgtggt catcgccgac ctgcacttca cgtaaatcga caggcgtaag 840  
 caggtgatgg actcgcgagg ccactacagc cggccggagt tgctcagcct cttaatagac 900  
 cgcaccccta ccgcgcactt tcacgaacgc gcttcgcacc ccacgacagg agctgagcaa 960  
 ggctccgagg atgtgttcga ggctaacatt taa 993

<210> 18  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 18  
 Met Arg Val Val Lys Ala Ala Ala Val Gln Leu Ser Pro Val Leu Tyr  
 1 5 10 15  
 Ser Arg Glu Gly Thr Val Glu Lys Val Val Arg Lys Ile His Glu Leu  
 20 25 30  
 Ala Glu Glu Gly Val Glu Phe Val Thr Phe Pro Glu Thr Val Val Pro  
 35 40 45  
 Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Thr Pro Leu Gln Gln Ile Phe  
 50 55 60  
 Gly Thr Glu Tyr Leu Arg Leu Leu Asp Gln Ala Val Thr Val Pro Ser  
 65 70 75 80  
 Ala Ala Thr Asp Ala Ile Gly Glu Ala Ala Arg Phe Ala Gly Val Val  
 85 90 95



Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Leu Tyr Asn Thr  
 100 105 110  
 Gln Leu Leu Phe Asp Ala Asp Gly Ser Leu Ile Gln Arg Arg Arg Lys  
 115 120 125  
 Ile Thr Pro Thr His Tyr Glu Arg Met Ile Trp Gly Gln Gly Asp Gly  
 130 135 140  
 Ser Gly Leu Arg Ala Val Asp Ser Lys Ala Gly Arg Ile Gly Gln Leu  
 145 150 155 160  
 Ala Cys Trp Glu His Asn Asn Pro Leu Ala Arg Tyr Ala Leu Ile Ala  
 165 170 175  
 Asp Gly Glu Gln Ile His Ser Ala Met Tyr Pro Gly Ser Met Phe Gly  
 180 185 190  
 Asp Ser Phe Ala Lys Lys Thr Glu Ile Asn Ile Arg Gln His Ala Leu  
 195 200 205  
 Glu Ser Ala Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Gly Asp  
 210 215 220  
 Gln Gln Ala Gln Ile Met Lys Asp Thr Gly Cys Ser Ile Gly Pro Ile  
 225 230 235 240  
 Ser Gly Gly Cys Phe Thr Thr Ile Val Ala Pro Asp Gly Ser Leu Ile  
 245 250 255  
 Gly Glu Pro Leu Arg Ser Gly Glu Gly Val Val Ile Ala Asp Leu Asp  
 260 265 270  
 Phe Thr Leu Ile Asp Arg Arg Lys Gln Val Met Asp Ser Arg Gly His  
 275 280 285  
 Tyr Ser Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Thr Pro Thr  
 290 295 300  
 Ala His Phe His Glu Arg Ala Ser His Pro Thr Thr Gly Ala Glu Gln  
 305 310 315 320  
 Gly Ser Glu Asp Val Phe Glu Ala Asn Ile  
 325 330

<210> 19  
 <211> 1050  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 19  
 atggaaagca acttccttgc cgcagcagtg caagcagaac cggtttactt caatgctttt 60  
 cagacggccg aaaaggccgc gtcattgatt gacgatgccg gtcggcaggg ggctcgctta 120  
 gtgacatttc ccgaaacgtg gctgcccggt taccggtact ggatctggct tggtgccccc 180  
 gcctggggaa tgcattcat tcatcctaaag taccatcaaa actcgccggt tgcaggagga 240  
 ccagaggaac agatcctttg tcaggcggcc cgccgcaacg ggatttttgt cgtcatggga 300  
 ctcagcgaga aaatcggggc aagcctctac atggcgcagt ggttcatcag tccagacggc 360  
 aaagtggtcg ctgcgccgacg caaattgaag cctactcacg tcgaacgttc ggtcttcggg 420  
 gaaggggatg gttccgacat tgtcgttctt gatacacccc ttggaaaggt cgggggcctt 480  
 tgctgctggg agcacatgca gccactttcg aagtacgcca tgtactcgca aggcgagcag 540  
 atccatgctg cttcttggcc gagtgtagc gtctatcgcg ataaaattta cgttctgggg 600  
 ccggagctga acggtgccgc caatcagatg tatgcggcag aaggtcagtg tttcgtcctg 660  
 gcatcctggg caacggtttc acaagcggct atcgatcttt tttgcgacac gcccgacaag 720  
 gccgcgctca tgaaaattgg tgggtggtttt tcccagatct atgggccaga cgggtgcccc 780  
 ctggcgaagc cgttgccgga ggacgtcgaa ggattggtga ccgctgagat tgacttcaat 840  
 gccatcacgc gcgtgaaagc agcggcggac cccgtagggc actatagccg gcccgatgta 900  
 ttccgcctgt tgttcaatcg tacgcgcaa gaacgcgtgg tttctgtcaa cacgtttgtg 960  
 ccaggtgtca cccagcgaac cgccaagaat gggtcggcgg acgaattggt cggtcacccg 1020  
 gagaacgctg tcgcccgggc tgcagagtaa 1050

<210> 20  
 <211> 349



WO 03/097810

PCT/US03/15712

<212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 20

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Ser | Asn | Phe | Leu | Ala | Ala | Ala | Val | Gln | Ala | Glu | Pro | Val | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Phe | Asn | Ala | Phe | Gln | Thr | Ala | Glu | Lys | Ala | Ala | Ser | Leu | Ile | Asp | Asp |
|     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Gly | Arg | Gln | Gly | Ala | Arg | Leu | Val | Thr | Phe | Pro | Glu | Thr | Trp | Leu |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Pro | Gly | Tyr | Pro | Tyr | Trp | Ile | Trp | Leu | Gly | Ala | Pro | Ala | Trp | Gly | Met |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| His | His | Phe | Ile | Leu | Lys | Tyr | His | Gln | Asn | Ser | Pro | Val | Ala | Gly | Gly |
| 65  |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     |     | 80  |
| Pro | Glu | Glu | Gln | Ile | Leu | Cys | Gln | Ala | Ala | Arg | Arg | Asn | Gly | Ile | Phe |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Val | Val | Met | Gly | Leu | Ser | Glu | Lys | Ile | Gly | Ala | Ser | Leu | Tyr | Met | Ala |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Gln | Trp | Phe | Ile | Ser | Pro | Asp | Gly | Lys | Val | Val | Ala | Arg | Arg | Arg | Lys |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Leu | Lys | Pro | Thr | His | Val | Glu | Arg | Ser | Val | Phe | Gly | Glu | Gly | Asp | Gly |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Ser | Asp | Ile | Val | Val | Leu | Asp | Thr | Pro | Leu | Gly | Lys | Val | Gly | Gly | Leu |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Cys | Cys | Trp | Glu | His | Met | Gln | Pro | Leu | Ser | Lys | Tyr | Ala | Met | Tyr | Ser |
|     |     |     |     | 165 |     |     |     | 170 |     |     |     |     |     | 175 |     |
| Gln | Gly | Glu | Gln | Ile | His | Ala | Ala | Ser | Trp | Pro | Ser | Val | Ser | Val | Tyr |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Arg | Asp | Lys | Ile | Tyr | Val | Leu | Gly | Pro | Glu | Leu | Asn | Gly | Ala | Ala | Asn |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Gln | Met | Tyr | Ala | Ala | Glu | Gly | Gln | Cys | Phe | Val | Leu | Ala | Ser | Trp | Ala |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Thr | Val | Ser | Gln | Ala | Ala | Ile | Asp | Leu | Phe | Cys | Asp | Thr | Pro | Asp | Lys |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Ala | Ala | Leu | Met | Lys | Ile | Gly | Gly | Gly | Phe | Ser | Gln | Ile | Tyr | Gly | Pro |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Asp | Gly | Cys | Pro | Leu | Ala | Lys | Pro | Leu | Pro | Glu | Asp | Val | Glu | Gly | Leu |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Val | Thr | Ala | Glu | Ile | Asp | Phe | Asn | Ala | Ile | Thr | Arg | Val | Lys | Ala | Ala |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Ala | Asp | Pro | Val | Gly | His | Tyr | Ser | Arg | Pro | Asp | Val | Phe | Arg | Leu | Leu |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Phe | Asn | Arg | Thr | Arg | Gln | Glu | Arg | Val | Val | Ser | Val | Asn | Thr | Phe | Val |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Pro | Gly | Val | Thr | Gln | Arg | Thr | Ala | Lys | Asn | Gly | Ser | Ala | Asp | Glu | Leu |
|     |     |     |     | 325 |     |     |     | 330 |     |     |     |     |     | 335 |     |
| Val | Gly | His | Pro | Glu | Asn | Ala | Val | Ala | Arg | Ala | Ala | Glu |     |     |     |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     |     |     |     |

<210> 21  
 <211> 1065  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 21



```

atggcactag aacatccgaa gtacgtggcg gccgtgggtc aggccgcgcc cgaattcctg      60
aatctagaca gagggatcga aaagacgacg gcattgatcg acgaagcggg acagaaaggg      120
gcgccctga ttgcatttcc ggaaacctgg ctgccgggct atccgtttca tgtctggctc      180
ggtcctccc catgggctgct tggctcagga ttctgtccagc gctatttcga caactcgatg      240
acgtacgata gtcctcaggg cgctgcactg agggacgctg ccgcgcgcaa cgggatcacg      300
gtgggtattgg gcttgtcgga gcgatgcggc ggcagcctct atatcgcgca atggatcatc      360
ggcccggatg gcgcgacggg cgccacgcgc cgcaaattgc ggccgactca tatcgagcgc      420
accgttttctg gcgatggcga cggcagcgat ctggcagtac acgatctcaa catcggccgc      480
cttggcgcac tgtgtctgct ggagcacatt cagccgctga ccaagtacgc gatgtatgcg      540
cagcacgaac aggtgcacgt cgcggcctgg ccgagcttct ccatgtatga attcgcgccc      600
gcgctcggtc acgaggtgaa caacgcagtc agccgcgtct atgccgttga gggatcgtgc      660
ttcgtgctcg cgccgtgcgc ggtcatcagc gagcaaattg tgcacatgtt gtgcgacacg      720
gcagacaagc gcgcgatgat acgtgccggc ggccgggcac cagtggcggt cgggccggac      780
ggcgaagctc tggctcgagaa actgccggaa aatgaggaag gcttgcgtct ggtcgatata      840
gatctcggtc gcatctcgct tgcgaaggct gcggccgacc ccgtcgggtca ctacgcgcgc      900
cccgatgtct tgcggctctg gttcgacaag caaccgcggc ggtgcgtcga acatgccggc      960
gagaacgacg cgtcgcgcag gtcgcacggg tcgtccgggt cacaatcgcc ggcgcaggat     1020
gggccggcga acgacatggt agaccgtcag gaaaacgtcg attga                       1065

```

<210> 22  
 <211> 354  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 22  
 Met Ala Leu Glu His Pro Lys Tyr Val Ala Ala Val Val Gln Ala Ala  
 1 5 10 15  
 Pro Glu Phe Leu Asn Leu Asp Arg Gly Ile Glu Lys Thr Ile Ala Leu  
 20 25 30  
 Ile Asp Glu Ala Gly Gln Lys Gly Ala Ala Leu Ile Ala Phe Pro Glu  
 35 40 45  
 Thr Trp Leu Pro Gly Tyr Pro Phe His Val Trp Leu Gly Pro Pro Ala  
 50 55 60  
 Trp Ala Leu Gly Ser Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Met  
 65 70 75 80  
 Thr Tyr Asp Ser Pro Gln Ala Ala Ala Leu Arg Asp Ala Ala Ala Arg  
 85 90 95  
 Asn Gly Ile Thr Val Val Leu Gly Leu Ser Glu Arg Cys Gly Gly Ser  
 100 105 110  
 Leu Tyr Ile Ala Gln Trp Ile Ile Gly Pro Asp Gly Ala Thr Val Ala  
 115 120 125  
 Thr Arg Arg Lys Leu Arg Pro Thr His Ile Glu Arg Thr Val Phe Gly  
 130 135 140  
 Asp Gly Asp Gly Ser Asp Leu Ala Val His Asp Leu Asn Ile Gly Arg  
 145 150 155 160  
 Leu Gly Ala Leu Cys Cys Trp Glu His Ile Gln Pro Leu Thr Lys Tyr  
 165 170 175  
 Ala Met Tyr Ala Gln His Glu Gln Val His Val Ala Ala Trp Pro Ser  
 180 185 190  
 Phe Ser Met Tyr Glu Phe Ala Pro Ala Leu Gly His Glu Val Asn Asn  
 195 200 205  
 Ala Val Ser Arg Val Tyr Ala Val Glu Gly Ser Cys Phe Val Leu Ala  
 210 215 220  
 Pro Cys Ala Val Ile Ser Glu Gln Met Val Asp Met Leu Cys Asp Thr  
 225 230 235 240  
 Ala Asp Lys Arg Ala Met Ile Arg Ala Gly Gly Gly His Ala Val Ala  
 245 250 255  
 Phe Gly Pro Asp Gly Glu Ala Leu Val Glu Lys Leu Pro Glu Asn Glu



WO 03/097810

PCT/US03/15712

```

                260                265                270
Glu Gly Leu Leu Leu Val Asp Ile Asp Leu Gly Arg Ile Ser Leu Ala
                275                280                285
Lys Ala Ala Ala Asp Pro Val Gly His Tyr Ala Arg Pro Asp Val Leu
                290                295                300
Arg Leu Trp Phe Asp Lys Gln Pro Arg Arg Cys Val Glu His Ala Gly
305                310                315                320
Glu Asn Asp Ala Ser Arg Arg Ser His Gly Ser Ser Gly Ser Gln Ser
                325                330                335
Pro Ala Gln Asp Gly Pro Ala Asn Asp Met Val Asp Arg Gln Glu Asn
                340                345                350
Val Asp

```

<210> 23  
 <211> 1005  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 23
atgaccgtta tcaaagcagc cgccattcaa atcagccccg tgctctacag cggggcgggg      60
acagtcgaga aagttgtag gaaggtaga gagctcgggg ccaaaggtgt ccgattcgct      120
acctttcccg aaaccatcat accgtactac ccgtacttct cgttcggttca gtcggcggttc      180
gacatgaagc ttgggagtga acatcagcgg ctgctcgacg aatcagtcac aattccttcg      240
tccgagacgg acgcgatcgc ccaggccgcc aaggaagcgg gcatggtggt gtccgtcggg      300
gtcaatgagc gcgatgggcg atccatctac aacactcaac ttctgttcga cgctgatggc      360
acgctcattc agcgtaggcg aaagatcacc ccgacctatc acgagcgcac gatttggggg      420
caaggcgatg gatccggcct acgcgcggtc gatagcgccg tgggcccggat cggccagctt      480
gcctgctggg agcactacct tcccctggcg cggtagccc tcatcgcgga cggagagcaa      540
atccactcgg caatgtatcc aggctcggtc gctggtcgcg tatttgccga gcagatagag      600
gttagtatcc gccagcacgc gcttgagtca gcctgcttcg tcgtcaacgc gaccggatgg      660
cttagcgccg agcagcaagc tcaaatagtg aaggataccg gatgcgtcgt tggaccaatc      720
tccggtggct gctttacggc gattgttgat ccggagggtc ggatcatggg ggcgccactc      780
aaggcaggtg agggggaggt catcgagat ctcgattttg cgcagattga tttccgcaag      840
cgtgtgatgg atacgcgagg gcactacagc cgccccgaac ttctaagcct cacgatcgac      900
cgagtcagc accatcacat gactgagcga ggcgccgatc accgtgtaga ccacgcaaag      960
ccaacgggtca ccgcagagca gtcggccgct gagccggcgg aatga      1005

```

<210> 24  
 <211> 334  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 24
Met Thr Val Ile Lys Ala Ala Ala Ile Gln Ile Ser Pro Val Leu Tyr
  1          5          10          15
Ser Arg Ala Gly Thr Val Glu Lys Val Val Arg Lys Val Arg Glu Leu
          20          25          30
Gly Ala Lys Gly Val Arg Phe Ala Thr Phe Pro Glu Thr Ile Ile Pro
          35          40          45
Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Ser Ala Phe Asp Met Lys Leu
          50          55          60
Gly Ser Glu His Gln Arg Leu Leu Asp Glu Ser Val Thr Ile Pro Ser
          65          70          75          80
Ser Glu Thr Asp Ala Ile Ala Gln Ala Ala Lys Glu Ala Gly Met Val

```



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |  |  |
| Val | Ser | Val | Gly | Val | Asn | Glu | Arg | Asp | Gly | Arg | Ser | Ile | Tyr | Asn | Thr |  |  |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |  |
| Gln | Leu | Leu | Phe | Asp | Ala | Asp | Gly | Thr | Leu | Ile | Gln | Arg | Arg | Arg | Lys |  |  |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |  |
| Ile | Thr | Pro | Thr | Tyr | His | Glu | Arg | Met | Ile | Trp | Gly | Gln | Gly | Asp | Gly |  |  |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |  |
| Ser | Gly | Leu | Arg | Ala | Val | Asp | Ser | Ala | Val | Gly | Arg | Ile | Gly | Gln | Leu |  |  |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |  |  |
| Ala | Cys | Trp | Glu | His | Tyr | Leu | Pro | Leu | Ala | Arg | Tyr | Ala | Leu | Ile | Ala |  |  |
|     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |  |  |
| Asp | Gly | Glu | Gln | Ile | His | Ser | Ala | Met | Tyr | Pro | Gly | Ser | Phe | Ala | Gly |  |  |
|     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |  |  |
| Pro | Leu | Phe | Ala | Glu | Gln | Ile | Glu | Val | Ser | Ile | Arg | Gln | His | Ala | Leu |  |  |
|     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |  |  |
| Glu | Ser | Ala | Cys | Phe | Val | Val | Asn | Ala | Thr | Gly | Trp | Leu | Ser | Ala | Glu |  |  |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |  |
| Gln | Gln | Ala | Gln | Ile | Val | Lys | Asp | Thr | Gly | Cys | Val | Val | Gly | Pro | Ile |  |  |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |  |
| Ser | Gly | Gly | Cys | Phe | Thr | Ala | Ile | Val | Asp | Pro | Glu | Gly | Arg | Ile | Met |  |  |
|     |     |     | 245 |     |     |     |     | 250 |     |     |     |     |     | 255 |     |  |  |
| Gly | Ala | Pro | Leu | Lys | Ala | Gly | Glu | Gly | Glu | Val | Ile | Ala | Asp | Leu | Asp |  |  |
|     |     | 260 |     |     |     | 265 |     |     |     |     |     |     | 270 |     |     |  |  |
| Phe | Ala | Gln | Ile | Asp | Phe | Arg | Lys | Arg | Val | Met | Asp | Thr | Arg | Gly | His |  |  |
|     | 275 |     |     |     |     | 280 |     |     |     |     |     | 285 |     |     |     |  |  |
| Tyr | Ser | Arg | Pro | Glu | Leu | Leu | Ser | Leu | Thr | Ile | Asp | Arg | Ser | Gln | His |  |  |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |  |  |
| His | His | Met | Thr | Glu | Arg | Gly | Ala | Asp | His | Arg | Val | Asp | His | Ala | Lys |  |  |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |  |
| Pro | Thr | Val | Thr | Ala | Glu | Gln | Ser | Ala | Val | Glu | Pro | Ala | Glu |     |     |  |  |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     |     |     |  |  |

<210> 25  
 <211> 939  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 25  
 gtgtcatcaa ctatcaaagt cgccattatt caggccgctc ctgcttacta cgacctgcag 60  
 gcgtcgctgg caaaggccgc cagtctgatc cgcgaggcgg cacgcggcgg cgcgcaattc 120  
 gtcgcgttcg gggagacatg gctgccgggc tatccgatgt ggctggattg gtgtcctggc 180  
 gcgatcatct gggataaccc cgccaccaa accgtcttcg cgcgccctcca tgaaaacagc 240  
 gtcgccgttc ccggcaggga aacggcattt ctgcgccgacc ttgcgatgtc gttaagcatc 300  
 gtattatgca tcggcgctcaa tgagaagggt atgaatgggc cgggacacgg cacgctctac 360  
 aacacgctcc tgacgtttga tgcaacgggt gaaatcatca atcatcatcg caagttgatg 420  
 ccaacctatg gcgagagatt ggtatggggg ccgggacgac cagttggcgt gcaagcgggt 480  
 gatagtacgg tcgggacgat cgggcggtg atctgttggg agcactggat gccgctgcca 540  
 cgccaactca tgcacaacag cggcgagcag attcacgtct gcgcatggcc gggcgtgcac 600  
 gaaatgcacc agatcgcgag ccgtcattat gcattcgagg gccgctgctt tgtgctggcc 660  
 gccggattga tcatgccgcg gttcgacctg ccagcgaaac tcgaatttcc gccgaactg 720  
 gccgacaagc gcgactatct cctaataaac ggcggcagcg ccatcatcaa gcccaatggc 780  
 aatatctcg ccgggcccgt ttatgacgaa gagactattc tctgcgccga ccttgacctg 840  
 actgagaaca tcaaggagca gatgacgctg gacgtgacag ggcattatgc gcgagcggaa 900  
 ctgtttgact tgaatgtggt gcggcgccgg aatgcgtag 939

<210> 26  
 <211> 312  
 <212> PRT



WO 03/097810

PCT/US03/15712

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 26

```

Val Ser Ser Thr Ile Lys Val Ala Ile Ile Gln Ala Ala Pro Ala Tyr
 1          5          10          15
Tyr Asp Leu Gln Ala Ser Leu Ala Lys Ala Ala Ser Leu Ile Arg Glu
          20          25          30
Ala Ala Arg Gly Gly Ala Gln Phe Val Ala Phe Gly Glu Thr Trp Leu
          35          40          45
Pro Gly Tyr Pro Met Trp Leu Asp Trp Cys Pro Gly Ala Ile Ile Trp
          50          55          60
Asp Asn Pro Ala Thr Lys Thr Val Phe Ala Arg Leu His Glu Asn Ser
65          70          75          80
Val Ala Val Pro Gly Arg Glu Thr Ala Phe Leu Ala Asp Leu Ala Met
          85          90          95
Ser Leu Ser Ile Val Leu Cys Ile Gly Val Asn Glu Lys Val Met Asn
          100          105          110
Gly Pro Gly His Gly Thr Leu Tyr Asn Thr Leu Leu Thr Phe Asp Ala
          115          120          125
Thr Gly Glu Ile Ile Asn His His Arg Lys Leu Met Pro Thr Tyr Gly
          130          135          140
Glu Arg Leu Val Trp Gly Pro Gly Asp Ala Val Gly Val Gln Ala Val
145          150          155          160
Asp Ser Thr Val Gly Arg Ile Gly Gly Leu Ile Cys Trp Glu His Trp
          165          170          175
Met Pro Leu Pro Arg Gln Leu Met His Asn Ser Gly Glu Gln Ile His
          180          185          190
Val Cys Ala Trp Pro Gly Val His Glu Met His Gln Ile Ala Ser Arg
          195          200          205
His Tyr Ala Phe Glu Gly Arg Cys Phe Val Leu Ala Ala Gly Leu Ile
          210          215          220
Met Pro Ala Phe Asp Leu Pro Ser Glu Leu Glu Phe Pro Pro Glu Leu
225          230          235          240
Ala Asp Lys Arg Asp Tyr Leu Leu Met Asn Gly Gly Ser Ala Ile Ile
          245          250          255
Lys Pro Asn Gly Lys Tyr Leu Ala Gly Pro Val Tyr Asp Glu Glu Thr
          260          265          270
Ile Leu Cys Ala Asp Leu Asp Leu Thr Glu Asn Ile Lys Glu Gln Met
          275          280          285
Thr Leu Asp Val Thr Gly His Tyr Ala Arg Ala Glu Leu Phe Asp Leu
          290          295          300
Asn Val Val Arg Arg Arg Asn Ala
305          310

```

&lt;210&gt; 27

&lt;211&gt; 1056

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 27

```

atgccaaccc ccagcgatca tttcaaaatc gccgctgttc aggcctcgcc cgtgtttctg      60
gaccgggagg ccactgtgga aaaggcctgc cggttgatcg ccgaagccgc aaagcagggc      120
gccgcctca tcgtctttcc ggaatctttc atcccgacct acccggattg ggtgtgggcc      180
gttcccccggaagggaaag aatcctgaac cagctgtatt ctgaattcct ggccaatgcc      240
gtcgatgttc ccggcgcggc gaccgaacaa cttgcccagg ctgcacgaat ggccggcgcc      300

```



|             |            |             |             |            |            |      |
|-------------|------------|-------------|-------------|------------|------------|------|
| tatgtgatta  | tgggcgtcac | cgaaagagac  | acctcggcca  | gcggggccag | cctctacaac | 360  |
| accctgctct  | acttcagccc | cgaaggcatc  | ctaattgggca | aacaccggaa | gctgggtccc | 420  |
| acggggggcg  | aacggctggg | ctgggcctac  | ggagacggca  | gcacgctgga | ggtctacgac | 480  |
| actccgctgg  | gaaagatcgg | cgggctgac   | tgctgggaga  | actacatgcc | cctggcccgg | 540  |
| tacacgatgt  | acgcctgggg | cacccagatt  | tacatcgccg  | ccacctggga | ccgcggggaa | 600  |
| ccgtggctct  | ccaccctgcg | gcataatcgcc | aaggaaggaa  | gggtctacgt | catcgggtgc | 660  |
| tgcatacgccc | tgcgccaggg | ggatatcccg  | gaccgggttcg | agtacaaggg | aaaattttat | 720  |
| tccgggtccc  | gggagtggat | caatgagggc  | gacagcgcca  | tcgtgaaccc | ggacggggaa | 780  |
| ttcatcgccg  | ggccggtgcg | gatgaaggag  | gagatcctgt  | atgccgagat | agacccccgg | 840  |
| cagatcgggg  | gccccaaagt | gatgctcgat  | gtggccggtc  | attacgcccg | gccggatatc | 900  |
| ttcgagctca  | tcgtccaccg | gaatccccac  | ccgatgatca  | aaatcgccga | agacaggggc | 960  |
| acggggatcg  | cctcaagttt | gattcgcccc  | cgccctaacc  | ttcccccatc | aagggggagg | 1020 |
| aaatcggcaa  | gaagcaaacg | caagcccaaa  | aaatga      |            |            | 1056 |

&lt;210&gt; 28

&lt;211&gt; 351

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 28

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Pro | Thr | Pro | Ser | Asp | His | Phe | Lys | Ile | Ala | Ala | Val | Gln | Ala | Ser |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Pro | Val | Phe | Leu | Asp | Arg | Glu | Ala | Thr | Val | Glu | Lys | Ala | Cys | Arg | Leu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ile | Ala | Glu | Ala | Ala | Lys | Gln | Gly | Ala | Arg | Leu | Ile | Val | Phe | Pro | Glu |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Ser | Phe | Ile | Pro | Thr | Tyr | Pro | Asp | Trp | Val | Trp | Ala | Val | Pro | Pro | Gly |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Arg | Glu | Arg | Ile | Leu | Asn | Gln | Leu | Tyr | Ser | Glu | Phe | Leu | Ala | Asn | Ala |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Val | Asp | Val | Pro | Gly | Ala | Ala | Thr | Glu | Gln | Leu | Ala | Gln | Ala | Ala | Arg |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Met | Ala | Gly | Ala | Tyr | Val | Ile | Met | Gly | Val | Thr | Glu | Arg | Asp | Thr | Ser |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ala | Ser | Gly | Ala | Ser | Leu | Tyr | Asn | Thr | Leu | Leu | Tyr | Phe | Ser | Pro | Glu |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Gly | Ile | Leu | Met | Gly | Lys | His | Arg | Lys | Leu | Val | Pro | Thr | Gly | Gly | Glu |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Arg | Leu | Val | Trp | Ala | Tyr | Gly | Asp | Gly | Ser | Thr | Leu | Glu | Val | Tyr | Asp |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| Thr | Pro | Leu | Gly | Lys | Ile | Gly | Gly | Leu | Ile | Cys | Trp | Glu | Asn | Tyr | Met |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Pro | Leu | Ala | Arg | Tyr | Thr | Met | Tyr | Ala | Trp | Gly | Thr | Gln | Ile | Tyr | Ile |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Ala | Ala | Thr | Trp | Asp | Arg | Gly | Glu | Pro | Trp | Leu | Ser | Thr | Leu | Arg | His |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Ile | Ala | Lys | Glu | Gly | Arg | Val | Tyr | Val | Ile | Gly | Cys | Cys | Ile | Ala | Leu |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Arg | Gln | Gly | Asp | Ile | Pro | Asp | Arg | Phe | Glu | Tyr | Lys | Gly | Lys | Phe | Tyr |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
| Ser | Gly | Ser | Arg | Glu | Trp | Ile | Asn | Glu | Gly | Asp | Ser | Ala | Ile | Val | Asn |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Pro | Asp | Gly | Glu | Phe | Ile | Ala | Gly | Pro | Val | Arg | Met | Lys | Glu | Glu | Ile |
|     |     |     | 260 |     |     |     | 265 |     |     |     |     |     | 270 |     |     |
| Leu | Tyr | Ala | Glu | Ile | Asp | Pro | Arg | Gln | Met | Arg | Gly | Pro | Lys | Trp | Met |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Leu | Asp | Val | Ala | Gly | His | Tyr | Ala | Arg | Pro | Asp | Ile | Phe | Glu | Leu | Ile |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |



Val His Arg Asn Pro His Pro Met Ile Lys Ile Ala Glu Asp Arg Gly  
 305 310 315 320  
 Thr Gly Ile Ala Ser Ser Leu Ile Arg Pro Arg Pro Asn Leu Pro Pro  
 325 330 335  
 Ser Arg Gly Arg Lys Ser Ala Arg Ser Lys Arg Lys Pro Lys Lys  
 340 345 350

<210> 29  
 <211> 1017  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 29  
 atgggcatcg aacatacgaa atacaagggtg gcggtggtgc aggcggcgcc ggcctggctc 60  
 gacctcgagg cctcgatcgg caagtccatc ggcctaataca aggaggccgc ggacaagggc 120  
 gccaaagctga tcgcctttcc ggaggccttc atccccgggt acccctggta tatctggatg 180  
 gactcgccgg cctgggcgat cggccgcggc ttctgtccagc gctatttcga caattcgctc 240  
 tcctacgaca gtccccaggc cgagcggctg cgtgatgccg tgcgccaggc caagctcacc 300  
 gccgtgatcg gcctgtccga acgcgacggc ggcagccttt acctggcgca atgggttgatc 360  
 gggcccgacg gcgaaaccat tgccaagcgc cgcaagctgc ggccgacca tgccgagcgc 420  
 accgtctatg gcgaaggcga cggcagcgat ctggccgtac atgcccggcc cgacatcggt 480  
 cgcttgggcg cgctgtgctg ctgggagcat cttcagccgt tgtcgaagta cgcaatgtac 540  
 gccagaacg agcagggtcca cgtcgctgcc tggccgagct tctcgctcta cgatcccttc 600  
 gcccgcggcg tcggcgccga ggtcaacaac gctgcctcgc gcgtctatgc ggtggagggc 660  
 tcctgcttcg tgctcgcgcc ttgcgcgacg gtgtcgcagg ccatgatcga cgaactctgc 720  
 gatcgcccg ataagcatgc gctgctgcat gccggcggag gctttgccgc gatctacggc 780  
 cccgacggca gccagatcgg cgagaagctg gcgccggatc aggagggtct gctgatcgcc 840  
 gagattgatc tgggcgccat cgggtgttgcc aagaacgcgg cagatcccgc cggtcattat 900  
 tcacggccgg atgtgacgcg gttgctgctc aacaagaagc ggtaccagcg cgtcgagcaa 960  
 ttgacctgac ccgccgacat ggtcgagccc gcggacatag gcgcggcgcc gagctga 1017

<210> 30  
 <211> 338  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 30  
 Met Gly Ile Glu His Thr Lys Tyr Lys Val Ala Val Val Gln Ala Ala  
 1 5 10 15  
 Pro Ala Trp Leu Asp Leu Glu Ala Ser Ile Gly Lys Ser Ile Gly Leu  
 20 25 30  
 Ile Lys Glu Ala Ala Asp Lys Gly Ala Lys Leu Ile Ala Phe Pro Glu  
 35 40 45  
 Ala Phe Ile Pro Gly Tyr Pro Trp Tyr Ile Trp Met Asp Ser Pro Ala  
 50 55 60  
 Trp Ala Ile Gly Arg Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu  
 65 70 75 80  
 Ser Tyr Asp Ser Pro Gln Ala Glu Arg Leu Arg Asp Ala Val Arg Gln  
 85 90 95  
 Ala Lys Leu Thr Ala Val Ile Gly Leu Ser Glu Arg Asp Gly Gly Ser  
 100 105 110  
 Leu Tyr Leu Ala Gln Trp Leu Ile Gly Pro Asp Gly Glu Thr Ile Ala  
 115 120 125  
 Lys Arg Arg Lys Leu Arg Pro Thr His Ala Glu Arg Thr Val Tyr Gly  
 130 135 140



Glu Gly Asp Gly Ser Asp Leu Ala Val His Ala Arg Pro Asp Ile Gly  
 145 150 155 160  
 Arg Leu Gly Ala Leu Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys  
 165 170 175  
 Tyr Ala Met Tyr Ala Gln Asn Glu Gln Val His Val Ala Ala Trp Pro  
 180 185 190  
 Ser Phe Ser Leu Tyr Asp Pro Phe Ala Pro Ala Leu Gly Ala Glu Val  
 195 200 205  
 Asn Asn Ala Ala Ser Arg Val Tyr Ala Val Glu Gly Ser Cys Phe Val  
 210 215 220  
 Leu Ala Pro Cys Ala Thr Val Ser Gln Ala Met Ile Asp Glu Leu Cys  
 225 230 235 240  
 Asp Arg Pro Asp Lys His Ala Leu Leu His Ala Gly Gly Gly Phe Ala  
 245 250 255  
 Ala Ile Tyr Gly Pro Asp Gly Ser Gln Ile Gly Glu Lys Leu Ala Pro  
 260 265 270  
 Asp Gln Glu Gly Leu Leu Ile Ala Glu Ile Asp Leu Gly Ala Ile Gly  
 275 280 285  
 Val Ala Lys Asn Ala Ala Asp Pro Ala Gly His Tyr Ser Arg Pro Asp  
 290 295 300  
 Val Thr Arg Leu Leu Leu Asn Lys Lys Arg Tyr Gln Arg Val Glu Gln  
 305 310 315 320  
 Phe Ala Leu Pro Ala Asp Met Val Glu Pro Ala Asp Ile Gly Ala Ala  
 325 330 335  
 Ala Ser

<210> 31  
 <211> 933  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 31  
 atgaccagaa tagccattat tcagcgcaccg cccgtgctgc tcgatcgaag cgccaccatt 60  
 gcccgggccg tgcaatcggc cgccgaagcg gcagcgcgaag gcgcgaccct gattgtcttg 120  
 cccgaatcgt acatccctgg ctatccctca tggatctggc ggctcgcgcc tggcaaagac 180  
 ggcgcgatcg tgggccagtt gcatgcgcgc ttgctggcca atgcgggtcga cctgagcagc 240  
 actgacctcg atgcgcttct tgaagcggcc cgtcagcacg gcgtgaccat tgtttgccgc 300  
 atgaacgagt gcgaacggcg tcgcggcgcc ggcacctgtg acaacacggg ggtcgtgatc 360  
 ggaccggacg gcgtcatgct caaccggcat cgcaaattga tgccgaccaa tcccagcgc 420  
 atggtgcatg gctttggcga tgcattccga ctgaaagcag ttgatacgcc tgccggcccg 480  
 ctgggcacgc tgatctgctg ggagagctac atgccgctgg cagcctatgc cctgtacgag 540  
 caaggcatcg agatctacat cgcaccaact tatgacagtg gtgacggctg gatcagcacc 600  
 atgcgccaca ttgcactcga agggcgctgc tgggtgattg gcagcggcac ggtcctgaaa 660  
 ggcagtgata ttccggacga tttcccggaa cgggcacgcc tggtccctga tccggatgag 720  
 tggatcaacg atggtgattc ggtagttatc gatccgcagg gaaagatcgt tgccgggtccg 780  
 atgcgtaggg aagcaggcat tctatacgcc gatatcgacg tcgcgcgcgt agcaccatca 840  
 cgccgcacgc tggatgtcgc ggggcattac gcgcgtccgg acgtcttcga gcttcgggta 900  
 caccaggcac cgggggcacg agtaagtaat tga 933

<210> 32  
 <211> 310  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample



&lt;400&gt; 32

```

Met Thr Arg Ile Ala Ile Ile Gln Arg Pro Pro Val Leu Leu Asp Arg
 1           5           10           15
Ser Ala Thr Ile Ala Arg Ala Val Gln Ser Val Ala Glu Ala Ala Ala
          20           25           30
Gln Gly Ala Thr Leu Ile Val Leu Pro Glu Ser Tyr Ile Pro Gly Tyr
          35           40           45
Pro Ser Trp Ile Trp Arg Leu Ala Pro Gly Lys Asp Gly Ala Ile Val
          50           55           60
Gly Gln Leu His Ala Arg Leu Leu Ala Asn Ala Val Asp Leu Ser Ser
65          70           75           80
Thr Asp Leu Asp Ala Leu Leu Glu Ala Ala Arg Gln His Gly Val Thr
          85           90           95
Ile Val Cys Gly Met Asn Glu Cys Glu Arg Arg Arg Gly Gly Gly Thr
          100          105          110
Leu Tyr Asn Thr Val Val Val Ile Gly Pro Asp Gly Val Met Leu Asn
          115          120          125
Arg His Arg Lys Leu Met Pro Thr Asn Pro Glu Arg Met Val His Gly
          130          135          140
Phe Gly Asp Ala Ser Gly Leu Lys Ala Val Asp Thr Pro Ala Gly Arg
145          150          155          160
Leu Gly Thr Leu Ile Cys Trp Glu Ser Tyr Met Pro Leu Ala Arg Tyr
          165          170          175
Ala Leu Tyr Glu Gln Gly Ile Glu Ile Tyr Ile Ala Pro Thr Tyr Asp
          180          185          190
Ser Gly Asp Gly Trp Ile Ser Thr Met Arg His Ile Ala Leu Glu Gly
          195          200          205
Arg Cys Trp Val Ile Gly Ser Gly Thr Val Leu Lys Gly Ser Asp Ile
          210          215          220
Pro Asp Asp Phe Pro Glu Arg Ala Arg Leu Phe Pro Asp Pro Asp Glu
225          230          235          240
Trp Ile Asn Asp Gly Asp Ser Val Val Ile Asp Pro Gln Gly Lys Ile
          245          250          255
Val Ala Gly Pro Met Arg Arg Glu Ala Gly Ile Leu Tyr Ala Asp Ile
          260          265          270
Asp Val Ala Arg Val Ala Pro Ser Arg Arg Thr Leu Asp Val Ala Gly
          275          280          285
His Tyr Ala Arg Pro Asp Val Phe Glu Leu Arg Val His Gln Ala Pro
          290          295          300
Gly Ala Arg Val Ser Asn
305          310

```

&lt;210&gt; 33

&lt;211&gt; 1026

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 33

```

atgttaagtc ccgtgacgca gtatcgcgcc gccgcggtgc aggcggcgcc atcttttctc      60
gatctcgacc gcaccgtcga gaagacgac gcgatcatcg agcaggcggc cgagcaggat      120
gtgcgccctga tcgcgtttcc ggaaacctgg attcccggtc atccgctctg gatctggctc      180
ggctcgccgg cctggggcat gcgcttcgtg cagcgctatt tcgagaactc gctggtgogc      240
ggcagcaaac agtggaacgc gatcgccgat gcggcgcggc gccaccgcat gaccgtcgtc      300
gtcggcttca gcgagcgcgc gggaggcagc ctctacatgg gccaggcgat cttcggcccc      360
gaaggcgagc tcatcgcggc gcgccggaag ctcaagccga cacacgccga gcgaacggtg      420
ttcggcgagg gcgacggcag ccacttggcc gtttacgaga cgggcgttgg tcgcatcggc      480
gccctctgct gctgggagca catccagccg ctctcgaaat acgcgatgta tgcggccaac      540
gaacagggtgc atgtggcctc gtggccgtgc ttcagccttt atcgcggcac ggcctatgcg      600

```



```

ctcggggccgg aggtgaacac cgccgcgagc caggtctacg cggtcgaggg cggctgctac 660
gtgctggcct cctgtctcgt cgtgacaccc gagatcctga aggtgctgat cgacacgccc 720
gacaaggagc cgttgctgct cgccggcggg gggttctcga tgatcttcgg ccccgacggc 780
cgcgcgctcg cccagccgct gccggagacc gaagaggggc tcgtcacggc cgagatcgat 840
ctcggcgcgga tcgcgctcgc caaggccgcg gccgatcccg ccggccatta cgcgcgggccc 900
gacgtgacgc ggttggttgc gaacccgcgc cccgcggcgc gcgtcgaagc gctgggtccg 960
cgcttcgagg tcgtgcagag cgagcaggcc gagccgcca cgcaaccggc cgaagcggcg 1020
gattga 1026

```

&lt;210&gt; 34

&lt;211&gt; 341

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 34

```

Met Leu Ser Pro Val Thr Gln Tyr Arg Ala Ala Ala Val Gln Ala Ala
1          5          10          15
Pro Ser Phe Leu Asp Leu Asp Arg Thr Val Glu Lys Thr Ile Ala Ile
20          25          30
Ile Glu Gln Ala Ala Glu Gln Asp Val Arg Leu Ile Ala Phe Pro Glu
35          40          45
Thr Trp Ile Pro Gly Tyr Pro Leu Trp Ile Trp Leu Gly Ser Pro Ala
50          55          60
Trp Gly Met Arg Phe Val Gln Arg Tyr Phe Glu Asn Ser Leu Val Arg
65          70          75          80
Gly Ser Lys Gln Trp Asn Ala Ile Ala Asp Ala Ala Arg Arg His Arg
85          90          95
Met Thr Val Val Val Gly Phe Ser Glu Arg Ala Gly Gly Ser Leu Tyr
100          105          110
Met Gly Gln Ala Ile Phe Gly Pro Glu Gly Glu Leu Ile Ala Ala Arg
115          120          125
Arg Lys Leu Lys Pro Thr His Ala Glu Arg Thr Val Phe Gly Glu Gly
130          135          140
Asp Gly Ser His Leu Ala Val Tyr Glu Thr Gly Val Gly Arg Ile Gly
145          150          155          160
Ala Leu Cys Cys Trp Glu His Ile Gln Pro Leu Ser Lys Tyr Ala Met
165          170          175
Tyr Ala Ala Asn Glu Gln Val His Val Ala Ser Trp Pro Cys Phe Ser
180          185          190
Leu Tyr Arg Gly Met Ala Tyr Ala Leu Gly Pro Glu Val Asn Thr Ala
195          200          205
Ala Ser Gln Val Tyr Ala Val Glu Gly Gly Cys Tyr Val Leu Ala Ser
210          215          220
Cys Leu Val Val Thr Pro Glu Ile Leu Lys Val Leu Ile Asp Thr Pro
225          230          235          240
Asp Lys Glu Pro Leu Leu Leu Ala Gly Gly Phe Ser Met Ile Phe
245          250          255
Gly Pro Asp Gly Arg Ala Leu Ala Gln Pro Leu Pro Glu Thr Glu Glu
260          265          270
Gly Leu Val Thr Ala Glu Ile Asp Leu Gly Ala Ile Ala Leu Ala Lys
275          280          285
Ala Ala Ala Asp Pro Ala Gly His Tyr Ala Arg Pro Asp Val Thr Arg
290          295          300
Leu Leu Leu Asn Pro Arg Pro Ala Ala Arg Val Glu Ala Leu Gly Pro
305          310          315          320
Arg Phe Glu Val Val Gln Ser Glu Gln Ala Glu Pro Pro Thr Gln Pro
325          330          335
Ala Glu Ala Ala Asp

```



340

<210> 35  
 <211> 942  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 35  
 atgacagctc taaaaattgc tgctgttcaa atgtgcgccg aattgggagc tacagatcga 60  
 aacctgagtg cagctggatc attcgtgcgc gacgcatttc gcgaagggtg ccagtgggta 120  
 atcctcccag agttttttac ctccgcaatg gcattcgcac cttcgatggc gcaagcttgg 180  
 ttgccactgg aaggaaaggc gctagcgatg atgcgcgacc ttgcgcgatc attcgatggg 240  
 gttgtttggag gctcatatgt tgccagagag gggaacgact gcgtaaatgc ctttcttctc 300  
 gtctttccgg atggaagcta ctaccggcat gacaaagata ttccaacaat gtgggagaaac 360  
 tgttactaca tcggcggcgt cgacgatggg gtgctggaaa caccaattgg tgcgggtggga 420  
 gttgcactgt gttgggagtt catccgaaca caaacgccc gaagactgaa ggatcgcggt 480  
 caattagtgg ttggcggtag ttgctggtgg gattttccga tgcctgtacc tgaacgatat 540  
 ctgaggctga ccaggcatat ctccaggaac tttgagcgcg atgctccggc gcggttggcc 600  
 agtatgttgg gtgtgcctgt tgtacacgct tcccatgctg gggattttac tgctgtcacc 660  
 ccaggcaatg aaacgaagaa ttaccgatcc aactatctgg gagagacca gatcgtcgat 720  
 gccaatggaa atgtgttgaa gcgaatgaca gtggctgatg gtgagggtta cgtcattgct 780  
 gacgttcaat tgggggccat atcaaccggt cgaacttcga tccccgacac cttctggacc 840  
 tgcaagctaa cgccaggggc acaacaggct tgggatgaac aaaatgcttt tgggtgtggc 900  
 tactatgaga acgtcacacg caaacaccta atcggctgat ga 942

<210> 36  
 <211> 313  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 36  
 Met Thr Ala Leu Lys Ile Ala Ala Val Gln Met Cys Ala Glu Leu Gly  
 1 5 10 15  
 Ala Thr Asp Arg Asn Leu Ser Ala Ala Gly Ser Phe Val Arg Asp Ala  
 20 25 30  
 Phe Arg Glu Gly Ala Gln Trp Val Ile Leu Pro Glu Phe Phe Thr Ser  
 35 40 45  
 Ala Met Ala Phe Ala Pro Ser Met Ala Gln Ala Trp Leu Pro Leu Glu  
 50 55 60  
 Gly Lys Ala Leu Ala Met Met Arg Ser Leu Ala Arg Gln Phe Asp Gly  
 65 70 75 80  
 Val Val Gly Gly Ser Tyr Val Ala Arg Glu Gly Asn Asp Cys Val Asn  
 85 90 95  
 Ala Phe Leu Leu Val Phe Pro Asp Gly Ser Tyr Tyr Arg His Asp Lys  
 100 105 110  
 Asp Ile Pro Thr Met Trp Glu Asn Cys Tyr Tyr Ile Gly Gly Val Asp  
 115 120 125  
 Asp Gly Val Leu Glu Thr Pro Ile Gly Ala Val Gly Val Ala Leu Cys  
 130 135 140  
 Trp Glu Phe Ile Arg Thr Gln Thr Ala Arg Arg Leu Lys Asp Arg Val  
 145 150 155 160  
 Gln Leu Val Val Gly Gly Thr Cys Trp Trp Asp Phe Pro Met Pro Val  
 165 170 175  
 Pro Glu Arg Tyr Leu Arg Leu Thr Arg His Ile Ser Arg Asn Phe Glu  
 180 185 190



```

Arg Asp Ala Pro Ala Arg Leu Ala Ser Met Leu Gly Val Pro Val Val
      195                200                205
His Ala Ser His Ala Gly Asp Phe Thr Ala Val Thr Pro Gly Asn Glu
      210                215                220
Thr Lys Asn Tyr Arg Ser Asn Tyr Leu Gly Glu Thr Gln Ile Val Asp
225      230                235                240
Ala Asn Gly Asn Val Leu Lys Arg Met Thr Val Ala Asp Gly Glu Gly
      245                250                255
Tyr Val Ile Ala Asp Val Gln Leu Gly Ala Ile Ser Thr Gly Arg Thr
      260                265                270
Ser Ile Pro Asp Thr Phe Trp Thr Cys Lys Leu Thr Pro Gly Ala Gln
      275                280                285
Gln Ala Trp Asp Glu Gln Asn Ala Phe Gly Cys Gly Tyr Tyr Glu Asn
      290                295                300
Val Thr Arg Lys His Leu Ile Gly Arg
305                310

```

<210> 37  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 37
atggattcag atatgacgga tactttttaag gcagccatta ttcagcacgc gcctgttttt 60
ttaaatctcg aagagagtct ggacaaagcc ggcagcctta tagaaaaggc tgccgatcaa 120
ggcgcgaaag tgatcgctt tcctgaaaca tggctgcccg gttatcccgt atggctcgac 180
tactctccaa aagcgggtct gtgggactat cagcctgcaa aatctctcta tcgtctgcta 240
gtcgataatt cagtcacctt acccgcaaaa cacctcgatc aactcctctc catagcgcaa 300
aagaccggcg catatgttgt aatgggggca cacgaacgag tgggtggaac actctataac 360
acgacgatct atgttgggat tgatgggaag gagtacaaac ttcatagaaa gctggtgccg 420
acctataccg aaagattgat ctggggggcg ggagacggca gcacattgag tgtgttgatg 480
acggattatg gcgttcttgg aggattgatc tgctgggagc actggatgcc tctggcaaga 540
gccgcaatgc atgccagata tgaaaccctt catgtggcgc aatggccggc tgtaaaagat 600
atccatcaga tagcaagcag acattatgct tttgaaggcc ggtgtttcgt gctcgcgga 660
ggctctgttc tgactcgaag agatataata gaaggattca actcactggc tcgcgccgat 720
agtgatgcat tggaacttct gaaagctatt tcgggagaag atagtgatct tattttgaat 780
gggggaagcg cgataattgc gccgaatgga gagtatcttg cgggcccggc ctttaatgaa 840
ccctccatta tttatgctga aattgatcct gcaactgataa gtgagggccca tcttacactg 900
gatacaagcg gacactactc gcgccctgac atttttcgtc tggagataaa cgatcaacct 960
caacatgatg taactttcag atcggggcat tag 993

```

<210> 38  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 38
Met Asp Ser Asp Met Thr Asp Thr Phe Lys Ala Ala Ile Ile Gln His
 1          5          10          15
Ala Pro Val Phe Leu Asn Leu Glu Glu Ser Leu Asp Lys Ala Gly Ser
      20          25          30
Leu Ile Glu Lys Ala Ala Asp Gln Gly Ala Lys Val Ile Ala Phe Pro
      35          40          45
Glu Thr Trp Leu Pro Gly Tyr Pro Val Trp Leu Asp Tyr Ser Pro Lys
      50          55          60

```



Ala Gly Leu Trp Asp Tyr Gln Pro Ala Lys Ser Leu Tyr Arg Leu Leu  
 65 70 75 80  
 Val Asp Asn Ser Val Thr Leu Pro Gly Lys His Leu Asp Gln Leu Leu  
 85 90 95  
 Ser Ile Ala Gln Lys Thr Gly Ala Tyr Val Val Met Gly Ala His Glu  
 100 105 110  
 Arg Val Gly Gly Thr Leu Tyr Asn Thr Thr Ile Tyr Val Gly Ile Asp  
 115 120 125  
 Gly Lys Glu Tyr Lys Leu His Arg Lys Leu Val Pro Thr Tyr Thr Glu  
 130 135 140  
 Arg Leu Ile Trp Gly Arg Gly Asp Gly Ser Thr Leu Ser Val Leu Met  
 145 150 155 160  
 Thr Asp Tyr Gly Val Leu Gly Gly Leu Ile Cys Trp Glu His Trp Met  
 165 170 175  
 Pro Leu Ala Arg Ala Ala Met His Ala Arg Tyr Glu Thr Leu His Val  
 180 185 190  
 Ala Gln Trp Pro Ala Val Lys Asp Ile His Gln Ile Ala Ser Arg His  
 195 200 205  
 Tyr Ala Phe Glu Gly Arg Cys Phe Val Leu Ala Ala Gly Ser Val Leu  
 210 215 220  
 Thr Arg Arg Asp Ile Ile Glu Gly Phe Asn Ser Leu Ala Arg Ala Asp  
 225 230 235 240  
 Ser Asp Ala Leu Glu Leu Leu Lys Ala Ile Ser Gly Glu Asp Ser Asp  
 245 250 255  
 Leu Ile Leu Asn Gly Gly Ser Ala Ile Ile Ala Pro Asn Gly Glu Tyr  
 260 265 270  
 Leu Ala Gly Pro Val Phe Asn Glu Pro Ser Ile Ile Tyr Ala Glu Ile  
 275 280 285  
 Asp Pro Ala Leu Ile Ser Glu Gly His Leu Thr Leu Asp Thr Ser Gly  
 290 295 300  
 His Tyr Ser Arg Pro Asp Ile Phe Arg Leu Glu Ile Asn Asp Gln Pro  
 305 310 315 320  
 Gln His Asp Val Thr Phe Arg Ser Gly His  
 325 330

&lt;210&gt; 39

&lt;211&gt; 1008

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 39

atgaaaaata tcaaaaactc agaaaaaagc agcacagtaa gagtcgctgc ggtacaaatc 60  
 agtccggtgt tgtacaaccg cgaagctacc gttcaaaaag tagtcaacaa aatccttgaa 120  
 ctaggaaaac aaggggtaca attcgccact tttccggaaa cgatagtgcc ttattatcct 180  
 tattttctctt ttattcaggc gccttatgcc atgggcaaag aacacctgcg cttgcttgaa 240  
 caatcagtta ctgttccgtc agccgcgacc gatgccataa gtgaggcggc aaaggaagcc 300  
 aatatggtag tgtctatttg tgtcaatgaa cgagacggtg gtaccattta caatacgcaa 360  
 ctcctttttg atgctgacgg aacattaatt cagcgcagac gtaaacttac accaacgtat 420  
 catgaaagaa tgatttgggg acaaggtgac gcttcagggtc ttcgtgccac agacagcgct 480  
 gttgggcgta tcgggcagtt ggcttggttg gaacattaca atccattggt ccgttatgct 540  
 ttgattgctg atggagaaca aatccattct gccatgtatc ccggatcatt tttagggtgcg 600  
 ttgcacggtg aacaaaccga aatcaatgta cgccaacacg ctttagaatc ggccagcttc 660  
 gtcgtagtgg ctaccggttg gttggatgcc gatcaacaag cacaattgc gaaagacacc 720  
 ggtggaccaa tcggaccaat ttcgggaggt tgttttacag ccgttatagg ccctgacgga 780  
 caactaatcg ggaagccct tacatcaggt gaaggggaag tgattgccga tattgatttg 840  
 gcacaaattg atgccgcaa aagattaatg gatgccagtg gtcactacaa ccgtcctgaa 900  
 ttgttgagct tgcatatcga tcacactccg actgctccta tgcataaaag agtagtttac 960  
 actgagccgg gattagcaaa aagacaaaat gaaaattcat caaatata 1008



WO 03/097810

PCT/US03/15712

<210> 40  
 <211> 335  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 40  
 Met Lys Asn Ile Lys Asn Ser Glu Lys Ser Ser Thr Val Arg Val Ala  
 1 5 10 15  
 Ala Val Gln Ile Ser Pro Val Leu Tyr Asn Arg Glu Ala Thr Val Gln  
 20 25 30  
 Lys Val Val Asn Lys Ile Leu Glu Leu Gly Lys Gln Gly Val Gln Phe  
 35 40 45  
 Ala Thr Phe Pro Glu Thr Ile Val Pro Tyr Tyr Pro Tyr Phe Ser Phe  
 50 55 60  
 Ile Gln Ala Pro Tyr Ala Met Gly Lys Glu His Leu Arg Leu Leu Glu  
 65 70 75 80  
 Gln Ser Val Thr Val Pro Ser Ala Ala Thr Asp Ala Ile Ser Glu Ala  
 85 90 95  
 Ala Lys Glu Ala Asn Met Val Val Ser Ile Gly Val Asn Glu Arg Asp  
 100 105 110  
 Gly Gly Thr Ile Tyr Asn Thr Gln Leu Leu Phe Asp Ala Asp Gly Thr  
 115 120 125  
 Leu Ile Gln Arg Arg Arg Lys Leu Thr Pro Thr Tyr His Glu Arg Met  
 130 135 140  
 Ile Trp Gly Gln Gly Asp Ala Ser Gly Leu Arg Ala Thr Asp Ser Ala  
 145 150 155 160  
 Val Gly Arg Ile Gly Gln Leu Ala Cys Trp Glu His Tyr Asn Pro Leu  
 165 170 175  
 Phe Arg Tyr Ala Leu Ile Ala Asp Gly Glu Gln Ile His Ser Ala Met  
 180 185 190  
 Tyr Pro Gly Ser Phe Leu Gly Ala Leu His Gly Glu Gln Thr Glu Ile  
 195 200 205  
 Asn Val Arg Gln His Ala Leu Glu Ser Ala Ser Phe Val Val Val Ala  
 210 215 220  
 Thr Gly Trp Leu Asp Ala Asp Gln Gln Ala Gln Ile Ala Lys Asp Thr  
 225 230 235 240  
 Gly Gly Pro Ile Gly Pro Ile Ser Gly Gly Cys Phe Thr Ala Val Ile  
 245 250 255  
 Gly Pro Asp Gly Gln Leu Ile Gly Glu Ala Leu Thr Ser Gly Glu Gly  
 260 265 270  
 Glu Val Ile Ala Asp Ile Asp Leu Ala Gln Ile Asp Ala Arg Lys Arg  
 275 280 285  
 Leu Met Asp Ala Ser Gly His Tyr Asn Arg Pro Glu Leu Leu Ser Leu  
 290 295 300  
 His Ile Asp His Thr Pro Thr Ala Pro Met His Glu Arg Val Val Tyr  
 305 310 315 320  
 Thr Glu Pro Gly Leu Ala Lys Arg Gln Asn Glu Asn Ser Ser Asn  
 325 330 335

<210> 41  
 <211> 966  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample



<400> 41  
 atgcccaatg agaataacat cgctacattc aaagttgccg cagtccaggc cacaccggtg 60  
 tttcttgatc gtgaagcaac catcgacaaa gcttgogggt tgattgccac tgccggcaat 120  
 gaaggagcgc gcctgattgt gtttccagaa gcgttcatcc caacctatcc tgaatgggtt 180  
 tgggggtattc cttccggtga gcaaggttta ctcaatgaac tctatgcaga gctgctcacc 240  
 aatgcggtca ctattcccag tgacgcgact gacaggctgt gcgaggccgc gcagcttgcg 300  
 aatgcctacg tagtgatggg aatgagcgaa cggaatgtcg aggcgagtgg cgcaagccta 360  
 tataatacgc tgttgatcat caatgcgcag ggggagattt tagggaaaca tcgaaagctg 420  
 gtgccaacgg gcggcgaacg cctgggtatgg gcgcagggtg atggcagcac gctgcaggtc 480  
 tacgatacgc cattgggaaa actcgggtgt ctcatttgct gggaaaatta tatgccgctg 540  
 gcacgctatg ctatgtatgc ctgggggaca caaatctatg tcgcggcaac gtgggatcga 600  
 ggccaaccct ggctttctac attacggcat atcgccaaag aaggcagggt atacgtgatt 660  
 ggttgctgta tcgcgatgcg aaaagacgat attccggata gttactccat gaagcagaaa 720  
 taccatgctg aatggatga atggattaat gttggcgaca gtgtgattgt caatcccgaa 780  
 ggacacttta tcgcagggcc tgtgcgcaag caagaagaaa ttctctacgc ggagatcgat 840  
 ccacgtatgg tgcaaggccc gaagtggatg ctcgatgtgg cggggcatta tgcgagacca 900  
 gatgtgttcc agttgacggt gcatacgaat gtgagagaga tgatgcgggt ggaagatgat 960  
 tcataa 966

<210> 42  
 <211> 321  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 42  
 Met Pro Asn Glu Asn Asn Ile Ala Thr Phe Lys Val Ala Ala Val Gln  
 1 5 10 15  
 Ala Thr Pro Val Phe Leu Asp Arg Glu Ala Thr Ile Asp Lys Ala Cys  
 20 25 30  
 Gly Leu Ile Ala Thr Ala Gly Asn Glu Gly Ala Arg Leu Ile Val Phe  
 35 40 45  
 Pro Glu Ala Phe Ile Pro Thr Tyr Pro Glu Trp Val Trp Gly Ile Pro  
 50 55 60  
 Ser Gly Glu Gln Gly Leu Leu Asn Glu Leu Tyr Ala Glu Leu Leu Thr  
 65 70 75 80  
 Asn Ala Val Thr Ile Pro Ser Asp Ala Thr Asp Arg Leu Cys Glu Ala  
 85 90 95  
 Ala Gln Leu Ala Asn Ala Tyr Val Val Met Gly Met Ser Glu Arg Asn  
 100 105 110  
 Val Glu Ala Ser Gly Ala Ser Leu Tyr Asn Thr Leu Leu Tyr Ile Asn  
 115 120 125  
 Ala Gln Gly Glu Ile Leu Gly Lys His Arg Lys Leu Val Pro Thr Gly  
 130 135 140  
 Gly Glu Arg Leu Val Trp Ala Gln Gly Asp Gly Ser Thr Leu Gln Val  
 145 150 155 160  
 Tyr Asp Thr Pro Leu Gly Lys Leu Gly Gly Leu Ile Cys Trp Glu Asn  
 165 170 175  
 Tyr Met Pro Leu Ala Arg Tyr Ala Met Tyr Ala Trp Gly Thr Gln Ile  
 180 185 190  
 Tyr Val Ala Ala Thr Trp Asp Arg Gly Gln Pro Trp Leu Ser Thr Leu  
 195 200 205  
 Arg His Ile Ala Lys Glu Gly Arg Val Tyr Val Ile Gly Cys Cys Ile  
 210 215 220  
 Ala Met Arg Lys Asp Asp Ile Pro Asp Ser Tyr Ser Met Lys Gln Lys  
 225 230 235 240  
 Tyr His Ala Glu Met Asp Glu Trp Ile Asn Val Gly Asp Ser Val Ile  
 245 250 255  
 Val Asn Pro Glu Gly His Phe Ile Ala Gly Pro Val Arg Lys Gln Glu



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |  |  |
| Glu | Ile | Leu | Tyr | Ala | Glu | Ile | Asp | Pro | Arg | Met | Val | Gln | Gly | Pro | Lys |  |  |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |  |  |
| Trp | Met | Leu | Asp | Val | Ala | Gly | His | Tyr | Ala | Arg | Pro | Asp | Val | Phe | Gln |  |  |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |  |  |
| Leu | Thr | Val | His | Thr | Asn | Val | Arg | Glu | Met | Met | Arg | Val | Glu | Asp | Asp |  |  |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |  |
| Ser |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |

<210> 43  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 43

|            |            |             |            |            |             |     |
|------------|------------|-------------|------------|------------|-------------|-----|
| atgagagtcg | ttaaagccgc | ggcgggtccaa | ctgaaacctg | tcctctatag | ccgagagggga | 60  |
| actgtcgaaa | acgtcgtccg | taaaatccac  | gagcttggac | agcaaggggt | acagttcgcg  | 120 |
| acgtttccag | agactgtggt | gccttactac  | ccgtactttt | cgatcgtgca | gtccggctat  | 180 |
| caaattcttg | gcggtggtga | gttcctgaag  | ctgcttgatc | agtcagtaac | cgtgccatct  | 240 |
| ctcgctacgg | aagcgatcgg | cgaggcttgc  | aggcaggcgg | gcgtcgttgt | ctccatcggc  | 300 |
| gtcaacgagc | gtgatggagg | aactctatac  | aacacgcaac | ttctctttga | tgccgacgga  | 360 |
| acattgattc | aaagacgacg | caagatcaca  | cccacccatt | acgagcgcat | ggtctggggc  | 420 |
| cagggcgatg | gctcaggttt | acggggccatt | gacagcaagg | tcgcgcgcat | tgggtcaactg | 480 |
| gcgtgttttg | agcactacaa | ccctctcgca  | cgttacgcga | tgatggccga | tggcgagcag  | 540 |
| atccattctg | cgatgttccc | cggctccatg  | ttcggcgata | atttttcaga | gaaggtggaa  | 600 |
| atcaacataa | ggcagcatgc | aatggagtct  | gggtgctttg | tcgtttgcgc | tactgcctgg  | 660 |
| ttggatgctg | accagcaggc | tcaaatcatg  | aaagacacgg | gatgtgagat | cggaccgatc  | 720 |
| tcaggagggt | gcttcacagc | gatcgcggca  | ccagatggaa | gccttatagg | tgaacccatc  | 780 |
| cgctcagggt | aaggcgtttg | tattgccgac  | ctcgatttca | aacttatcga | caagcggaag  | 840 |
| cacgtagtag | acacacgcgg | ccattatagc  | cggccagaat | tgctcagcct | cctgattgat  | 900 |
| cggacgccga | cggcccacat | acacgaaagg  | accgagcaac | cgagggcggc | catcgagaaa  | 960 |
| gagtcgcagg | atgttttcac | cgctgttgct  | ttaa       |            |             | 993 |

<210> 44  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 44

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Arg | Val | Val | Lys | Ala | Ala | Ala | Val | Gln | Leu | Lys | Pro | Val | Leu | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Ser | Arg | Glu | Gly | Thr | Val | Glu | Asn | Val | Val | Arg | Lys | Ile | His | Glu | Leu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Gly | Gln | Gln | Gly | Val | Gln | Phe | Ala | Thr | Phe | Pro | Glu | Thr | Val | Val | Pro |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Tyr | Tyr | Pro | Tyr | Phe | Ser | Ile | Val | Gln | Ser | Gly | Tyr | Gln | Ile | Leu | Gly |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Gly | Gly | Glu | Phe | Leu | Lys | Leu | Leu | Asp | Gln | Ser | Val | Thr | Val | Pro | Ser |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Leu | Ala | Thr | Glu | Ala | Ile | Gly | Glu | Ala | Cys | Arg | Gln | Ala | Gly | Val | Val |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Val | Ser | Ile | Gly | Val | Asn | Glu | Arg | Asp | Gly | Gly | Thr | Leu | Tyr | Asn | Thr |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |
| Gln | Leu | Leu | Phe | Asp | Ala | Asp | Gly | Thr | Leu | Ile | Gln | Arg | Arg | Arg | Lys |



Page 30



<223> Obtained from an environmental sample

<400> 46

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Lys | Pro | Pro | Thr | Ser | Phe | Arg | Val | Ala | Ala | Val | Gln | Ala | Cys | Pro |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Val | Tyr | Leu | Asp | Arg | Asp | Leu | Thr | Ile | Gly | Lys | Ala | Glu | Gly | Leu | Ile |
|     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Glu | Ala | Ala | Gly | Asn | Gly | Ala | Asn | Leu | Ile | Val | Phe | Pro | Glu | Ala |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Phe | Val | Pro | Gly | Tyr | Pro | Val | Trp | Val | Trp | Phe | Ile | Pro | Pro | Gly | Arg |
|     | 50  |     |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |
| Thr | Ala | Asp | Leu | Arg | Glu | Ala | Tyr | Ser | Val | Leu | His | Ala | Asn | Ser | Ile |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Ala | Val | Pro | Ser | Gln | Ser | Thr | Glu | Arg | Leu | Cys | Ala | Ala | Ala | Arg | Arg |
|     |     |     | 85  |     |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Ala | Gly | Val | Ala | Val | Ala | Ile | Gly | Val | Asn | Glu | Arg | Asn | Ser | Glu | Ala |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |
| Ser | Gly | Gly | Ser | Leu | Phe | Asn | Thr | Leu | Leu | Tyr | Ile | Gly | Pro | Asp | Gly |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Thr | Leu | Leu | Gly | Lys | His | Arg | Lys | Leu | Val | Pro | Thr | Gly | Gly | Glu | Arg |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Leu | Val | Trp | Ala | Ser | Gly | Asp | Gly | Ser | Asp | Leu | Ala | Val | Phe | Thr | Leu |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Pro | Phe | Ala | Arg | Val | Gly | Gly | Leu | Ile | Cys | Trp | Glu | Asn | Tyr | Met | Pro |
|     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Leu | Ala | Arg | Tyr | Ala | Leu | Ala | Ala | Trp | Gly | Ala | Gln | Ile | His | Val | Ala |
|     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Pro | Thr | Trp | Asp | Arg | Gly | Glu | Pro | Trp | Leu | Ser | Thr | Leu | Arg | His | Val |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Ala | Lys | Glu | Gly | Arg | Ala | Val | Thr | Ile | Gly | Cys | Cys | Gln | Ala | Val | Arg |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Lys | Glu | Asp | Ile | Pro | Asp | Gly | Leu | Ala | Phe | Lys | Ser | Arg | Tyr | Leu | Ala |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Asp | Val | Gly | Ala | Trp | Val | Asn | Pro | Gly | Gly | Ser | Val | Ile | Val | Asp | Pro |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Asp | Gly | Lys | Ile | Leu | Ala | Gly | Pro | Ala | Asn | Glu | Thr | Glu | Gly | Ile | Leu |
|     |     | 260 |     |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Tyr | Ala | Asp | Ile | Arg | Ala | Asp | Gln | Leu | Val | Gly | Pro | Arg | Trp | Gln | Leu |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Asp | Ile | Ala | Gly | His | Tyr | Ala | Arg | Pro | Asp | Val | Phe | Glu | Leu | Ile | Val |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| His | Arg | Arg | Ser | Thr | Pro | Met | Ile | Arg | Glu | Val | Ser | Ala | Pro | Arg | Arg |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Arg | Ala | Arg | Thr | Gly | Lys | Arg | Pro | Arg | Arg | Arg |     |     |     |     |     |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     |     |     |

<210> 47

<211> 1014

<212> DNA

<213> Unknown

<220>

<223> Obtained from an environmental sample

<400> 47

|             |            |            |            |            |             |     |
|-------------|------------|------------|------------|------------|-------------|-----|
| gtgaaagaag  | caatcaaagt | agcctgtgtg | caagcagctc | cagtctttct | cgacctggac  | 60  |
| gccacagtgg  | acaagaccgt | cgccctgatt | gaggaggcag | cccgtaacgg | cgcacgccta  | 120 |
| atcgccctttc | cagagacctg | gattccaggc | tacctatggt | tcctttggct | ggactcacca  | 180 |
| gcctggggga  | tgcaattcgt | gcgccgatac | cacgagaact | cactggctct | cgacagccct  | 240 |
| caggccaagc  | gcatcagtga | ggccgcccag | cgccgcccgt | tatacgctgc | gctaggggtac | 300 |
| agcgaacgcg  | tgagcgggaa | cctctacatg | gggcagtggc | tcattgacga | taagggcgaa  | 360 |



```

acagctgggc tgcgccgaaa gctgaaacca acccatgtag agcgaaccct cttcggtgaa 420
ggcgacggat catccctttc cacttttcgac acaccgttgg ggggtgctggg cggactctgc 480
tggtgggaac acttacaacc tcttttcgaaa tatgcgctct acgcacagaa cgaggaaata 540
cacttcgccg cctggcctag cttcagcatc taccgtcaag cgacagaagt ccttggaaca 600
gaagtaaagt tcgcagcttc tcggatctac gccgtggaag ggcagtgttt tggtctcgct 660
tcctgcgcgc tcgtctcgcc agagatgatc gaaatgctct gcactgacga aagcaagcac 720
agccttcttc aggccggcgg cgggtactcc cgcattatcg gtcccgatgg cagcgacct 780
gcgcgcccct tgggcgaaaa cgaggaaggt attctctatg ccactctgga ccctgccgct 840
cgaatctatg caaagaccgc agctgatcca gccgggcact actccagacc agacgtcact 900
cggctgctga tcaatcgag tgccaatcag ccagtcgtag aggttggaag ggaaatacct 960
gcatcggccc aaggctttga agttgaggcg gccccgggt acgaaggcga ttga 1014

```

&lt;210&gt; 48

&lt;211&gt; 337

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 48

```

Val Lys Glu Ala Ile Lys Val Ala Cys Val Gln Ala Ala Pro Val Phe
  1           5           10           15
Leu Asp Leu Asp Ala Thr Val Asp Lys Thr Val Ala Leu Ile Glu Glu
          20           25           30
Ala Ala Arg Asn Gly Ala Arg Leu Ile Ala Phe Pro Glu Thr Trp Ile
          35           40           45
Pro Gly Tyr Pro Trp Phe Leu Trp Leu Asp Ser Pro Ala Trp Gly Met
          50           55           60
Gln Phe Val Arg Arg Tyr His Glu Asn Ser Leu Val Leu Asp Ser Pro
          65           70           75           80
Gln Ala Lys Arg Ile Ser Glu Ala Ala Gln Arg Ala Gly Ile Tyr Val
          85           90           95
Ala Leu Gly Tyr Ser Glu Arg Val Ser Gly Thr Leu Tyr Met Gly Gln
          100          105          110
Trp Leu Ile Asp Asp Lys Gly Glu Thr Ala Gly Leu Arg Arg Lys Leu
          115          120          125
Lys Pro Thr His Val Glu Arg Thr Leu Phe Gly Glu Gly Asp Gly Ser
          130          135          140
Ser Leu Ser Thr Phe Asp Thr Pro Leu Gly Val Leu Gly Gly Leu Cys
          145          150          155          160
Cys Trp Glu His Leu Gln Pro Leu Ser Lys Tyr Ala Leu Tyr Ala Gln
          165          170          175
Asn Glu Glu Ile His Phe Ala Ala Trp Pro Ser Phe Ser Ile Tyr Arg
          180          185          190
Gln Ala Thr Glu Val Leu Gly Pro Glu Val Asn Val Ala Ala Ser Arg
          195          200          205
Ile Tyr Ala Val Glu Gly Gln Cys Phe Val Leu Ala Ser Cys Ala Leu
          210          215          220
Val Ser Pro Glu Met Ile Glu Met Leu Cys Thr Asp Glu Ser Lys His
          225          230          235          240
Ser Leu Leu Gln Ala Gly Gly Gly Tyr Ser Arg Ile Ile Gly Pro Asp
          245          250          255
Gly Ser Asp Leu Ala Arg Pro Leu Gly Glu Asn Glu Glu Gly Ile Leu
          260          265          270
Tyr Ala Thr Leu Asp Pro Ala Ala Arg Ile Tyr Ala Lys Thr Ala Ala
          275          280          285
Asp Pro Ala Gly His Tyr Ser Arg Pro Asp Val Thr Arg Leu Leu Ile
          290          295          300
Asn Arg Ser Ala Asn Gln Pro Val Val Glu Val Gly Arg Glu Ile Pro
          305          310          315          320

```







```

145          150          155          160
Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met Ala Asp Arg
          165          170          175
Glu Gln Ile His Val Ser Met Phe Pro Gly Ser Leu Val Gly Glu Ile
          180          185          190
Phe Ala Glu Gln Ile Glu Ala Thr Ile Arg His His Ala Leu Glu Ser
          195          200          205
Gly Cys Phe Val Val Asn Ala Thr Gly Trp Leu Thr Pro Glu Gln Gln
          210          215          220
Ala Gln Ile Val Lys Asp Thr Gly Gly Pro Ile Ala Ala Ile Ser Gly
225          230          235          240
Gly Cys Phe Thr Ala Ile Val Ser Pro Glu Gly Lys Leu Leu Gly Thr
          245          250          255
Pro Leu Arg Ser Asp Ser Gly Glu Gly Ala Cys Ile Ala Glu Leu Asp
          260          265          270
Phe Asn Leu Ile Asn Lys Arg Lys Arg Met Met Asp Ser Val Gly His
          275          280          285
Tyr Ser Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Lys Thr Pro Thr
          290          295          300
Ser His Thr His Pro Leu Lys Lys Pro Leu Ala Pro Ser Glu Lys Asn
305          310          315          320
Thr Pro Glu Asp Ile Ala Thr Gly Leu Thr Leu Val Thr Pro Val Ser
          325          330          335
Asn Ala Asn Leu Phe Ser Ala Ser Asn
          340          345

```

<210> 51  
 <211> 897  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 51
gtgaacgtcc gcgtcgcggt ggtgcaggcc acgccggccg tgctcgacgg gccggcgctcg      60
gtgcggaagg cctgccgcct gatcggcgag gccgcggccg gcggcgccc cctgatcgcc      120
ctgcccagag gcttcgtgcc catcatgccg cgctcctgct gggggcacca cttcgcgctg      180
atcgccctgc cgaagtcggc ggccctgcac cggcgcatct gggagaacgc cgtcgacgtc      240
ggcggcccgc tggcccgcga gctcggcgac gccgcgcgcc gcgcggacgc ctgggtggcc      300
atcggggtga acgagcgcga cgcccgcgg cggggcacgc tctggaacac gctgctctgg      360
ttcgcgcccg acgggagcct ggcccggcgc caccgcaagc tcgtgcccac catgcacgag      420
cgcacgttct gggggcaggg cgcgggacgc gacctcgagg cgctggccgc ggacttcggc      480
cgcttgggcg gcctgatctg ctgggagaaac ttcatgcccg ccgcgcgccg gcgcctgcac      540
cgggacgggg tcgacttcta cctggccccc acggcgagac accgggacat ctgggtcgcc      600
gcgatgcgca cgttcgcctt cgaggccggc gccttcgtcc tctcgccggt gcagtacctg      660
cggaccgccg acttcccgga ggacttcccg ctgcgcgagg agctcgccga ctgccccgag      720
gtccagttca ccggggggag cgtgatctgc gaccgtggg gcaacctcct ggcggggccg      780
gtccacgggg gcgaggagat cctctacgcc gactgcgata tcgacctcgt cctcgaggcc      840
cgacgggtgc tcgacacggc cggccactac gaccgcccgg acctcgccctc ggcctga      897

```

<210> 52  
 <211> 298  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 52
Val Asn Val Arg Val Ala Val Val Gln Ala Thr Pro Ala Val Leu Asp

```



|     |     |     |     |
|-----|-----|-----|-----|
| 1   | 5   | 10  | 15  |
| Gly | Pro | Ala | Ser |
| 20  | 25  | 30  |     |
| Ala | Gly | Gly | Ala |
| 35  | 40  | 45  |     |
| Met | Pro | Arg | Ser |
| 50  | 55  | 60  |     |
| Lys | Ser | Ala | Ala |
| 65  | 70  | 75  |     |
| Gly | Gly | Pro | Leu |
| 85  | 90  | 95  |     |
| Ala | Trp | Val | Ala |
| 100 | 105 | 110 |     |
| Thr | Leu | Trp | Asn |
| 115 | 120 | 125 |     |
| Arg | Arg | His | Arg |
| 130 | 135 | 140 |     |
| Gly | Gln | Gly | Ala |
| 145 | 150 | 155 |     |
| Arg | Leu | Gly | Gly |
| 165 | 170 | 175 |     |
| Arg | Arg | Leu | His |
| 180 | 185 | 190 |     |
| Asp | Asp | Arg | Asp |
| 195 | 200 | 205 |     |
| Ala | Gly | Ala | Phe |
| 210 | 215 | 220 |     |
| Phe | Pro | Glu | Asp |
| 225 | 230 | 235 |     |
| Val | Gln | Phe | Thr |
| 245 | 250 | 255 |     |
| Leu | Ala | Gly | Pro |
| 260 | 265 | 270 |     |
| Asp | Leu | Asp | Leu |
| 275 | 280 | 285 |     |
| His | Tyr | Asp | Arg |
| 290 | 295 |     |     |

<210> 53  
 <211> 954  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

|            |             |
|------------|-------------|
| <400> 53   |             |
| atggtcaatt | caaagagtca  |
| gaacgggagg | cgacgattga  |
| gcgaatctgg | tggtcttccc  |
| gttccggcag | gtgaaacaac  |
| ggtgaaattc | cggtccgggc  |
| tatgtcgcga | ttggcttgac  |
| actttgctct | ttctcgactc  |
| accggcgggc | agcgcttggt  |
| actaggtttg | gaaaaatggg  |
| tatgccttgt | atgcctgggg  |
| ccgtggctgt | caacgctgcg  |
| ggcatggccc | tgcgcaaagc  |
| cagaacgccg | atgagtggat  |
| ttcatcgccg | ggccgctgcg  |
|            | gattaaaatc  |
|            | caaagcttgc  |
|            | tgagtcattc  |
|            | gctcctgaac  |
|            | gacagagcgg  |
|            | agccggcggc  |
|            | ctgggctcag  |
|            | aggggttgatt |
|            | gacgcagatt  |
|            | gcatatcgcc  |
|            | cgacctgccc  |
|            | caatgtcggc  |
|            | cgagcaggaa  |
|            | gcggcggtgc  |
|            | cggttgattg  |
|            | gtcccggctt  |
|            | acgctctatg  |
|            | ctgagccagg  |
|            | atgttaggca  |
|            | ggtgatggca  |
|            | tgctgggaga  |
|            | tacatcgccg  |
|            | gcggaaggcc  |
|            | gaccgctttg  |
|            | gacagcgcca  |
|            | ggcatcctct  |
|            | aggctgcccc  |
|            | cggaggcagg  |
|            | atcccgattg  |
|            | ccgaactgct  |
|            | cagccaacct  |
|            | gtggggcgag  |
|            | agcatcgcaa  |
|            | gcactctggc  |
|            | attacatgcc  |
|            | ccacctggga  |
|            | gggttgttgt  |
|            | aactcaagca  |
|            | ttgttaaccc  |
|            | ggcatcctct  |
|            | atgctgagat  |
|            | ggcctttctg  |
|            | cgagcagggc  |
|            | ggctctgggc  |
|            | ggccaatgcc  |
|            | ggccgggggt  |
|            | cctgtacaat  |
|            | actgatcccc  |
|            | ggtgtacgag  |
|            | cctggcccgt  |
|            | tcgaggcgag  |
|            | cgctggctgt  |
|            | gcgattttac  |
|            | tgatggtgaa  |
|            | tgatctggcc  |



cagatgcgcg gccccaaatg gatgctcgac gtggccggcc attacgctcg cccggatgtg 900  
 tttgaactca tcgttcatcg ggaggcgcgg cccatgattg cgctaatttc atga 954

<210> 54  
 <211> 317  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 54  
 Met Val Asn Ser Lys Ser Gln Ile Lys Ile Ala Ala Val Gln Ala Ala  
 1 5 10 15  
 Pro Val Phe Leu Glu Arg Glu Ala Thr Ile Asp Lys Ala Cys Arg Leu  
 20 25 30  
 Ile Ala Glu Ala Gly Glu Gln Gly Ala Asn Leu Val Val Phe Pro Glu  
 35 40 45  
 Ser Phe Val Pro Ala Tyr Pro Asp Trp Val Trp Ala Val Pro Ala Gly  
 50 55 60  
 Glu Thr Thr Leu Leu Asn Thr Leu Tyr Ala Glu Leu Leu Ala Asn Ala  
 65 70 75 80  
 Val Glu Ile Pro Gly Pro Ala Thr Glu Arg Leu Ser Gln Ala Ala Asn  
 85 90 95  
 Leu Ala Gly Val Tyr Val Ala Ile Gly Leu Thr Glu Arg Asn Ile Glu  
 100 105 110  
 Ala Ser Gly Ala Ser Leu Tyr Asn Thr Leu Leu Phe Leu Asp Ser Ala  
 115 120 125  
 Gly Gly Met Leu Gly Lys His Arg Lys Leu Ile Pro Thr Gly Gly Glu  
 130 135 140  
 Arg Leu Val Trp Ala Gln Gly Asp Gly Ser Thr Leu Ala Val Tyr Glu  
 145 150 155 160  
 Thr Arg Phe Gly Lys Met Gly Gly Leu Ile Cys Trp Glu Asn Tyr Met  
 165 170 175  
 Pro Leu Ala Arg Tyr Ala Leu Tyr Ala Trp Gly Thr Gln Ile Tyr Ile  
 180 185 190  
 Ala Ala Thr Trp Asp Arg Gly Glu Pro Trp Leu Ser Thr Leu Arg His  
 195 200 205  
 Ile Ala Ala Glu Gly Arg Val Val Val Val Gly Cys Gly Met Ala Leu  
 210 215 220  
 Arg Lys Ala Asp Leu Pro Asp Arg Phe Glu Leu Lys Gln Arg Phe Tyr  
 225 230 235 240  
 Gln Asn Ala Asp Glu Trp Ile Asn Val Gly Asp Ser Ala Ile Val Asn  
 245 250 255  
 Pro Asp Gly Glu Phe Ile Ala Gly Pro Leu Arg Glu Gln Glu Gly Ile  
 260 265 270  
 Leu Tyr Ala Glu Ile Asp Leu Ala Gln Met Arg Gly Pro Lys Trp Met  
 275 280 285  
 Leu Asp Val Ala Gly His Tyr Ala Arg Pro Asp Val Phe Glu Leu Ile  
 290 295 300  
 Val His Arg Glu Ala Arg Pro Met Ile Ala Leu Ile Ser  
 305 310 315

<210> 55  
 <211> 1017  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample



&lt;400&gt; 55

```

atgggtatcg aacatccgaa gtacaagggtc gccgtgggtgc aggcagctcc cgcctggctc      60
gatctcgacg cgtcgatcga caagtcgacg gcgctgatcg aggaggcggc ccaaaaaggc      120
gccaagctga tcgcattccc cgaggccttc attcccggct atccctggca catctggatg      180
gactcgccgg cctggggcgat cggccgcggc tttgtgcagc gctattttga caattcgctc      240
gcctatgaca gcccgcaggc cgagaagctg cgcgcggcgg tgcgcaaggc aaagctcacc      300
gccgtgctcg ggctgtccga gcgcgacggc ggcagtctct atctggcgca atggttgatc      360
gggcccgatg gcgagaccat cgcaaaacgc cgcaagctgc ggccgacaca tgccgagcgc      420
acggtgtacg gcgagggcga cggcagcgat ctgcgagtc acaaccgtcc cgatatcggc      480
cgctcgggcg cgctctgctg ctgggagcat ttgcagccac tgtcgaaata cgcgatgtac      540
gcgcagaacg agcaggtgca tgtcgcggcc tggccgagct tttcgctcta cgatcccttt      600
gcggtggcgc tcggcgccga ggtgaacaac gcggcctcgc gcgtctatgc agtcgaaggc      660
tcttgcttcg tgctggcgcc atgcgccacc gtctcgcagg ccatgatcga cgagctctgc      720
gaccgaccgg acaagcatat gctgctgcat gtcggcggcg gttttgccgc gatctatggt      780
cctgacggca gccagatcgg cgacaagctc gcgcccgacc aggaagggtt gttgatcgcg      840
gagatcgacc ttggggccat tggcgctgcc aagaacgcgg ccgatcccgc cgggcattat      900
tcgcggcccg acgtgacgcg gctcctgctc aacaagaaac cgtacaagcg cgtcgagcag      960
ttctcgccac cggccgaggc ggtcgagccc acagatatcg cagcggcggc aagctga      1017

```

&lt;210&gt; 56

&lt;211&gt; 338

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 56

```

Met Gly Ile Glu His Pro Lys Tyr Lys Val Ala Val Val Gln Ala Ala
 1          5          10          15
Pro Ala Trp Leu Asp Leu Asp Ala Ser Ile Asp Lys Ser Ile Ala Leu
 20          25          30
Ile Glu Glu Ala Ala Gln Lys Gly Ala Lys Leu Ile Ala Phe Pro Glu
 35          40          45
Ala Phe Ile Pro Gly Tyr Pro Trp His Ile Trp Met Asp Ser Pro Ala
 50          55          60
Trp Ala Ile Gly Arg Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu
 65          70          75          80
Ala Tyr Asp Ser Pro Gln Ala Glu Lys Leu Arg Ala Ala Val Arg Lys
 85          90          95
Ala Lys Leu Thr Ala Val Leu Gly Leu Ser Glu Arg Asp Gly Gly Ser
100          105          110
Leu Tyr Leu Ala Gln Trp Leu Ile Gly Pro Asp Gly Glu Thr Ile Ala
115          120          125
Lys Arg Arg Lys Leu Arg Pro Thr His Ala Glu Arg Thr Val Tyr Gly
130          135          140
Glu Gly Asp Gly Ser Asp Leu Ala Val His Asn Arg Pro Asp Ile Gly
145          150          155          160
Arg Leu Gly Ala Leu Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys
165          170          175
Tyr Ala Met Tyr Ala Gln Asn Glu Gln Val His Val Ala Ala Trp Pro
180          185          190
Ser Phe Ser Leu Tyr Asp Pro Phe Ala Val Ala Leu Gly Ala Glu Val
195          200          205
Asn Asn Ala Ala Ser Arg Val Tyr Ala Val Glu Gly Ser Cys Phe Val
210          215          220
Leu Ala Pro Cys Ala Thr Val Ser Gln Ala Met Ile Asp Glu Leu Cys
225          230          235          240
Asp Arg Pro Asp Lys His Thr Leu Leu His Val Gly Gly Gly Phe Ala
245          250          255
Ala Ile Tyr Gly Pro Asp Gly Ser Gln Ile Gly Asp Lys Leu Ala Pro

```



WO 03/097810

PCT/US03/15712

```

                260                265                270
Asp Gln Glu Gly Leu Leu Ile Ala Glu Ile Asp Leu Gly Ala Ile Gly
                275                280                285
Val Ala Lys Asn Ala Ala Asp Pro Ala Gly His Tyr Ser Arg Pro Asp
                290                295                300
Val Thr Arg Leu Leu Leu Asn Lys Lys Pro Tyr Lys Arg Val Glu Gln
305                310                315                320
Phe Ser Pro Pro Ala Glu Ala Val Glu Pro Thr Asp Ile Ala Ala Ala
                325                330                335
Ala Ser

```

<210> 57  
 <211> 1014  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 57
gtgaaagaag caatcaaagt agcctgtgtg caagcagctc cagtctttct cgacctggac      60
gccacagtgg acaagaccgt cgccctgatt gaggaggcag cccgtaacgg cgcacgccta      120
atcgcccttc cagagacctg gattccaggc taccatggt tcctttggct ggactcacca      180
gcctggggga tgcaattcgt gcgccgatac cagcagaact cactggctct cgacagccct      240
caggccaagc gcatcagtga ggccgcccag cgcgccggtatacgtcgc gctaggggtac      300
agcgaacgcg tgagcggaac cctctacatg gggcagtggc tcattgacga taagggcgaa      360
acagctgggc tgcgccgaaa gctgaaacca acccatgtag agcgaaccct cttcggtgaa      420
ggcgacggat catccctttc cactttcgac acaccgttgg gggtgctggg cggactctgc      480
tgttgggaac acttacaacc tctttcgaaa tatgcgctct acgcacagaa cgaggaaata      540
cacttcgccg cctggcctag cttcagcatc taccgtcaag cgacagaagt ccttggacca      600
gaagtaaatg tcgcagcttc tcggatctac gccgtggaag ggcagtgttt tgttctcgct      660
tcctgcgcgc tcgtctcgcc agagatgatc gaaatgctct gcactgacga aagcaagcac      720
agccttcttc aggccggcgg cggtactcc cgcattatcg gtcccgatgg cagcgacct      780
gcgcgcccct tgggcgaaaa cgaggaaggt attctctatg ccactctgga ccctgccgct      840
cgaatctatg caaagaccgc agctgatcca gccgggcact actccagacc agacgtcact      900
cggctgctga tcaatcgtag tgccaatcag ccagtcgtag aggttggacg ggaaatacct      960
gcacggccc aaggctttga agttgaggcg gccccgggt acggaggcga ttga      1014

```

<210> 58  
 <211> 337  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 58
Val Lys Glu Ala Ile Lys Val Ala Cys Val Gln Ala Ala Pro Val Phe
  1                5                10                15
Leu Asp Leu Asp Ala Thr Val Asp Lys Thr Val Ala Leu Ile Glu Glu
                20                25                30
Ala Ala Arg Asn Gly Ala Arg Leu Ile Ala Phe Pro Glu Thr Trp Ile
                35                40                45
Pro Gly Tyr Pro Trp Phe Leu Trp Leu Asp Ser Pro Ala Trp Gly Met
                50                55                60
Gln Phe Val Arg Arg Tyr His Glu Asn Ser Leu Val Leu Asp Ser Pro
65                70                75                80
Gln Ala Lys Arg Ile Ser Glu Ala Ala Gln Arg Ala Gly Ile Tyr Val
                85                90                95
Ala Leu Gly Tyr Ser Glu Arg Val Ser Gly Thr Leu Tyr Met Gly Gln

```



|                       |     |
|-----------------------|-----|
| $\langle 210 \rangle$ | 60  |
| $\langle 211 \rangle$ | 328 |



WO 03/097810

PCT/US03/15712

<212> PRT  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 60  
Met Arg Asp Arg Asn Phe Lys Leu Ala Ala Ile Gln Ala Glu Pro Val  
1 5 10 15  
Phe Phe Asn Arg Arg Ala Ser Thr Glu Lys Ala Cys Arg Leu Ile Lys  
20 25 30  
Glu Ala Gly Ala Met Gly Ala Asp Ile Ala Gly Phe Ser Glu Thr Trp  
35 40 45  
Leu Pro Gly Tyr Pro Phe Phe Ile Trp Gly Ala Ser Ala Asp Pro Ser  
50 55 60  
Leu Leu Trp Lys Ala Ser Ala Glu Tyr Leu Ala Asn Ala Val Gln Ile  
65 70 75 80  
Pro Gly Pro Glu Thr Asp Gln Leu Cys Glu Ala Ala Lys Lys Ala Gly  
85 90 95  
Ile Asp Val Ala Ile Gly Val Val Glu Leu Asp Glu Phe Thr Lys Gly  
100 105 110  
Thr Ala Tyr Cys Thr Leu Leu Phe Ile Gly Lys Glu Gly Lys Ile Leu  
115 120 125  
Gly Lys His Arg Lys Leu Lys Pro Thr His Arg Glu Arg Thr Val Trp  
130 135 140  
Gly Glu Gly Asp Ala Thr Gly Leu Ser Val His Glu Arg Pro Tyr Gly  
145 150 155 160  
Arg Ile Ser Gly Leu Asn Cys Trp Glu His Asn Met Val Leu Pro Gly  
165 170 175  
Tyr Val Leu Met Ser Gln Gly Thr His Ile His Ile Ala Ala Trp Pro  
180 185 190  
Gly Ser Glu Gly Lys Ala Pro Pro Ala Pro Ser Pro Met Trp Glu Arg  
195 200 205  
Gln Leu Leu Leu Ser Arg Ala Phe Ala Ser Gln Ser Ala Ala Tyr Val  
210 215 220  
Ile Leu Val Gly Gly Leu Leu Asn Pro Gln Asn Ile Pro Ala Pro Tyr  
225 230 235 240  
Asp Glu Leu Ala Val Lys Tyr Arg Gly Asp Ser Phe Ile Ile Asp Pro  
245 250 255  
Arg Gly Glu Ile Ile Ala Gly Pro Ala Lys Gly Glu Thr Ile Leu Ile  
260 265 270  
Ala Glu Gly Ser Met Glu Gln Val Leu Ala Ala Lys Ser Ala Phe Asp  
275 280 285  
Val Ala Gly His Tyr Ser Arg Pro Asp Val Phe Gln Leu Cys Val Asn  
290 295 300  
Arg Lys Pro Tyr Arg Arg Val Arg Glu Thr Ser Glu Gln Asp Gln Pro  
305 310 315 320  
Ala Ser Glu Arg Glu Ser Glu Ser  
325

<210> 61  
<211> 966  
<212> DNA  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 61  
atgactcgat cttacccgaa tgacacactc acggttgggc ttgcgcaaatt tgctccagtc 60  
tggttgatc gtacagggac aatttcaaag atattagctc aagtccatgc ggcaaattgaa 120



```

gcgggctgtc atcttgtcgc gtttggcgaa ggtctcctcc ccggatatcc gttttggatt 180
gagcgaacaa atggtgcagt cttcaactcg cccacgcaga aagaaattca cgcgcattat 240
ctggatcagg ctgtccagat cgaagcaggt catcttgagg cgctttgcga agcagccaag 300
gaatatgaga tcgcaattgt cctgggatgc attgaacgtc cgcaagatcg tggagggcac 360
agtctgtatg caagccttgt atatattgat tcagacggca tcatccaatc tgtgcatcga 420
aagttaatgc caacatatga agaacggctc acctgggtcg caggtgacgg acatggatta 480
cgggtgcaca aattaggtgc ctttacgggt ggcggcctca actgttggga aaactggatg 540
cctttggcac gcgcggccat gtatggtcaa ggcgaggatt tgcattattgc catttggccc 600
ggcggtctccc acaatacgca agacattaca cgctttattg cactagaatc gcgttcctat 660
gttttatctg tgtcaggttt aatgcgctca ggcgattttc caaaagagac cccacatctt 720
gcatccatcc tggctaaagg tgaggatatt cttgccaacg gtggttcatg tatcgccggt 780
cctgacggca aatggatcgt tgagccgctt gtaggagaag agaagttaat tgttgcaacg 840
attgatcatt gtcgtgtgcg cgaagagcgt caaaattttg atccttcagg acattacagc 900
aggccagatg tattgcaact gaaaataaac aggcaacgcc agagtacaat ctcgtttgga 960
gagtaa 966

```

&lt;210&gt; 62

&lt;211&gt; 321

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 62

```

Met Thr Arg Ser Tyr Pro Asn Asp Thr Leu Thr Val Gly Leu Ala Gln
1      5      10      15
Ile Ala Pro Val Trp Leu Asp Arg Thr Gly Thr Ile Ser Lys Ile Leu
20      25      30
Ala Gln Val His Ala Ala Asn Glu Ala Gly Cys His Leu Val Ala Phe
35      40      45
Gly Glu Gly Leu Leu Pro Gly Tyr Pro Phe Trp Ile Glu Arg Thr Asn
50      55      60
Gly Ala Val Phe Asn Ser Pro Thr Gln Lys Glu Ile His Ala His Tyr
65      70      75      80
Leu Asp Gln Ala Val Gln Ile Glu Ala Gly His Leu Glu Ala Leu Cys
85      90      95
Glu Ala Ala Lys Glu Tyr Glu Ile Ala Ile Val Leu Gly Cys Ile Glu
100     105     110
Arg Pro Gln Asp Arg Gly Gly His Ser Leu Tyr Ala Ser Leu Val Tyr
115     120     125
Ile Asp Ser Asp Gly Ile Ile Gln Ser Val His Arg Lys Leu Met Pro
130     135     140
Thr Tyr Glu Glu Arg Leu Thr Trp Ser Pro Gly Asp Gly His Gly Leu
145     150     155     160
Arg Val His Lys Leu Gly Ala Phe Thr Val Gly Gly Leu Asn Cys Trp
165     170     175
Glu Asn Trp Met Pro Leu Ala Arg Ala Ala Met Tyr Gly Gln Gly Glu
180     185     190
Asp Leu His Ile Ala Ile Trp Pro Gly Gly Ser His Asn Thr Gln Asp
195     200     205
Ile Thr Arg Phe Ile Ala Leu Glu Ser Arg Ser Tyr Val Leu Ser Val
210     215     220
Ser Gly Leu Met Arg Ser Gly Asp Phe Pro Lys Glu Thr Pro His Leu
225     230     235     240
Ala Ser Ile Leu Ala Lys Gly Glu Asp Ile Leu Ala Asn Gly Gly Ser
245     250     255
Cys Ile Ala Gly Pro Asp Gly Lys Trp Ile Val Glu Pro Leu Val Gly
260     265     270
Glu Glu Lys Leu Ile Val Ala Thr Ile Asp His Cys Arg Val Arg Glu
275     280     285

```



WO 03/097810

PCT/US03/15712

Glu Arg Gln Asn Phe Asp Pro Ser Gly His Tyr Ser Arg Pro Asp Val  
 290 295 300  
 Leu Gln Leu Lys Ile Asn Arg Gln Arg Gln Ser Thr Ile Ser Phe Gly  
 305 310 315 320  
 Glu

<210> 63  
 <211> 978  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 63  
 atgcaagata gagtaccgat tgtacgagct gcggtatcc aggctgaacc catagtcctt 60  
 gattgtgacg cgaccgtgga aaaagcctgc cgattgatcg gtgaagcagc agaaaatggt 120  
 gcaaacctga tcgtgtttcc cgaagccttc attcccgttt atcccaatgc ggcgatctgg 180  
 ggctcgaggtc tggccacttt tggcggacag cgccagaaat acgtatggac gcgactatgg 240  
 aacaattcgg tggaaatccc tggtcgggcc accgacaggc tggcaaaggc agcacacgag 300  
 gctcgagcca ccgttgtcat gggattgaat gagcgcgcg tggataacaa cacgctttac 360  
 aacaccctgc tatttattgg gccagacggg cgcttgctgg gcaagcaccg taagctcatg 420  
 cccaccaatc acgaacggat gatctggggg atgggagatg ggagcaccct gcgggttttt 480  
 gatacaccct gtggaaaagt aggcgggtct atctgctggg aaaactacat gcctctggcg 540  
 cgttatgcac tctatggaca gggcgaacaa atccatgtcg cgccgactgc gcacgatggg 600  
 gagatcactc tgggtcaatgc acgcaatacc gcctatgagg gacgcttatt cgtcatctcc 660  
 gtgtgcatga tccttcgcaa gtccagcttt ccccatgatt ttgagctggg cgaggaattg 720  
 gcggaggcag atgacttcat aaaatcaggc ggcagcgcg tcgttgggcc agatggcgag 780  
 gtgctggcgg gtccattgtg gaatgaagag aatatactgt atgccgatct tgacttgaat 840  
 cgaattgtgg atgagagacg agtatttgat gtgacgggcc attattcacg tccagatgtt 900  
 ctacgactgc actttaatgc ttcccctcag aaaactattg aaagatatga gcaacctctc 960  
 gatccgtctg agggttaa 978

<210> 64  
 <211> 325  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 64  
 Met Gln Asp Arg Val Pro Ile Val Arg Ala Ala Ala Ile Gln Ala Glu  
 1 5 10 15  
 Pro Ile Val Leu Asp Cys Asp Ala Thr Val Glu Lys Ala Cys Arg Leu  
 20 25 30  
 Ile Gly Glu Ala Ala Glu Asn Gly Ala Asn Leu Ile Val Phe Pro Glu  
 35 40 45  
 Ala Phe Ile Pro Val Tyr Pro Asn Ala Ala Ile Trp Gly Arg Gly Leu  
 50 55 60  
 Ala Thr Phe Gly Gly Gln Arg Gln Lys Tyr Val Trp Thr Arg Leu Trp  
 65 70 75 80  
 Asn Asn Ser Val Glu Ile Pro Gly Pro Ala Thr Asp Arg Leu Ala Lys  
 85 90 95  
 Ala Ala His Glu Ala Arg Ala Thr Val Val Met Gly Leu Asn Glu Arg  
 100 105 110  
 Ala Val Asp Asn Asn Thr Leu Tyr Asn Thr Leu Leu Phe Ile Gly Pro  
 115 120 125  
 Asp Gly Arg Leu Leu Gly Lys His Arg Lys Leu Met Pro Thr Asn His  
 130 135 140



Glu Arg Met Ile Trp Gly Met Gly Asp Gly Ser Thr Leu Arg Val Phe  
 145 150 155 160  
 Asp Thr Pro Cys Gly Lys Val Gly Gly Leu Ile Cys Trp Glu Asn Tyr  
 165 170 175  
 Met Pro Leu Ala Arg Tyr Ala Leu Tyr Gly Gln Gly Glu Gln Ile His  
 180 185 190  
 Val Ala Pro Thr Ala His Asp Gly Glu Ile Thr Leu Val Asn Ala Arg  
 195 200 205  
 Asn Thr Ala Tyr Glu Gly Arg Leu Phe Val Ile Ser Val Cys Met Ile  
 210 215 220  
 Leu Arg Lys Ser Ser Phe Pro His Asp Phe Glu Leu Gly Glu Glu Leu  
 225 230 235 240  
 Ala Glu Ala Asp Asp Phe Ile Lys Ser Gly Gly Ser Ala Ile Val Gly  
 245 250 255  
 Pro Asp Gly Glu Val Leu Ala Gly Pro Leu Trp Asn Glu Glu Asn Ile  
 260 265 270  
 Leu Tyr Ala Asp Leu Asp Leu Asn Arg Ile Val Asp Glu Arg Arg Val  
 275 280 285  
 Phe Asp Val Thr Gly His Tyr Ser Arg Pro Asp Val Leu Arg Leu His  
 290 295 300  
 Phe Asn Ala Ser Pro Gln Lys Thr Ile Glu Arg Tyr Glu Gln Pro Leu  
 305 310 315 320  
 Asp Pro Ser Glu Gly  
 325

<210> 65  
 <211> 1002  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 65  
 atgccgaccc ccacttcaaa attcaaaatc ggcgccgtgc aggcacgcgc ggtttttctg 60  
 gaccgggaag ccactgcgca aaaagcctgc aaattgattg ccgaagcggg agggcagggc 120  
 ggcgggctga tcgtttttccc ggagtccttc attcccacct atcctgattg ggtctgggcg 180  
 gtcccgcgga gaagaggaaa agtggttaagc gaactttacg ccgagctgct ggccaatgcc 240  
 gtggaagtcc ccggggccggt caccgatcag ctgggtgaag cagcccaaaa aacgggcgcc 300  
 tatgtcgtca tgggcggtcac ggaaaaggac accgacgcaa gcggcgcgag cctttacaac 360  
 acgtcctct atttcaaccc cgcgggggac ctccctggga aacaccggaa gcttggtcct 420  
 accggcgggg agcggctggt ctgggcgcag ggcgacggca gcaccctgga agtgtacgac 480  
 actcccctgg gaaaaatcgg aggcctcatc tgctgggaaa actacatgcc cctcgcccgg 540  
 tacacgatgt atgcctgggg gaccagatt tatatcgcg ccacatggga ccagggggag 600  
 acgtggcttg ccaccctgcg gcatatcgct aaggaaggac ggggtgtacgt catcggtgc 660  
 tgcacgcgc tgcggcgga cgacatcccc gaccggctgg aatacaagaa gaagttctac 720  
 tcggggtcgc gggaatgat caatatgggg gacagcgcca tcgtgaaccc ggaaggcgaa 780  
 ttcatgtccg gcccgtgcg gatgaaggag gagatcctgt atgccgaggt ggacccctc 840  
 ctgatggcgg gatcgaaatg gatgctcgac gtcgcggggc attacgcgcg ccccgacgtc 900  
 tttgaactca tcgtccaccg ccagccccac ccgatgatcc gggtaatcga gaaagaggga 960  
 ggggcccggaa gaaccgggga cgagaagaag gaaaatgagt ga 1002

<210> 66  
 <211> 333  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 66



Met Pro Thr Pro Thr Ser Lys Phe Lys Ile Gly Ala Val Gln Ala Ser  
 1 5 10 15  
 Pro Val Phe Leu Asp Arg Glu Ala Thr Ala Gln Lys Ala Cys Lys Leu  
 20 25 30  
 Ile Ala Glu Ala Gly Gly Gln Gly Ala Arg Leu Ile Val Phe Pro Glu  
 35 40 45  
 Ser Phe Ile Pro Thr Tyr Pro Asp Trp Val Trp Ala Val Pro Pro Gly  
 50 55 60  
 Arg Gly Lys Val Leu Ser Glu Leu Tyr Ala Glu Leu Leu Ala Asn Ala  
 65 70 75 80  
 Val Glu Val Pro Gly Pro Val Thr Asp Gln Leu Gly Glu Ala Ala Gln  
 85 90 95  
 Lys Thr Gly Ala Tyr Val Val Met Gly Val Thr Glu Lys Asp Thr Asp  
 100 105 110  
 Ala Ser Gly Ala Ser Leu Tyr Asn Thr Leu Leu Tyr Phe Asn Pro Ala  
 115 120 125  
 Gly Asp Leu Leu Gly Lys His Arg Lys Leu Val Pro Thr Gly Gly Glu  
 130 135 140  
 Arg Leu Val Trp Ala Gln Gly Asp Gly Ser Thr Leu Glu Val Tyr Asp  
 145 150 155 160  
 Thr Pro Leu Gly Lys Ile Gly Gly Leu Ile Cys Trp Glu Asn Tyr Met  
 165 170 175  
 Pro Leu Ala Arg Tyr Thr Met Tyr Ala Trp Gly Thr Gln Ile Tyr Ile  
 180 185 190  
 Ala Ala Thr Trp Asp Gln Gly Glu Thr Trp Leu Ala Thr Leu Arg His  
 195 200 205  
 Ile Ala Lys Glu Gly Arg Val Tyr Val Ile Gly Cys Cys Ile Ala Leu  
 210 215 220  
 Arg Arg Asp Asp Ile Pro Asp Arg Leu Glu Tyr Lys Lys Lys Phe Tyr  
 225 230 235 240  
 Ser Gly Ser Arg Glu Trp Ile Asn Met Gly Asp Ser Ala Ile Val Asn  
 245 250 255  
 Pro Glu Gly Glu Phe Ile Ala Gly Pro Val Arg Met Lys Glu Glu Ile  
 260 265 270  
 Leu Tyr Ala Glu Val Asp Pro Leu Leu Met Ala Gly Ser Lys Trp Met  
 275 280 285  
 Leu Asp Val Ala Gly His Tyr Ala Arg Pro Asp Val Phe Glu Leu Ile  
 290 295 300  
 Val His Arg Gln Pro His Pro Met Ile Arg Val Ile Glu Lys Glu Gly  
 305 310 315 320  
 Gly Ala Gly Arg Thr Gly Asp Glu Lys Lys Glu Asn Glu  
 325 330

&lt;210&gt; 67

&lt;211&gt; 936

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 67

|            |             |             |            |            |            |     |
|------------|-------------|-------------|------------|------------|------------|-----|
| atgccccgtg | tggcgggtggt | ccagcgcgccg | ccggtgtttc | tgcaccgcgc | cgcgaccctc | 60  |
| gagaacgccg | tggcttcgct  | cgccgaggcc  | gcgtcgaacg | gggctcgcct | cgcggtcttt | 120 |
| ccggaagccc | tggttcccgg  | ctatccggcg  | tggatgtggc | ggctgcggcc | cgggcccgc  | 180 |
| atggcgctca | ccgagcggat  | tcacgcgcgc  | ttgcgggcga | actcggtag  | cctcgccgcc | 240 |
| gacgagctcg | cgccgctgcg  | cgaggcggcc  | cggcgccacg | agctcaccgt | agtgtgcggc | 300 |
| ctgcacgagc | gcgacgaggc  | gctcggcggc  | ggcacgctct | ataacaccgt | cgtcacgatc | 360 |
| ggcgccgacg | gcgcggtgct  | caaccgccac  | cggaagctga | tgcccaccaa | ccccgagcgc | 420 |
| atggtctggg | gctgcggcga  | tgccagcggg  | ctcaggacgg | tccccacca  | gtgcgggcgc | 480 |
| gtcggcgccc | tgatctgctg  | ggaaagctac  | atgccgcttg | cacgctacgc | gctgtacgcc | 540 |



```

caggggaatcg acctctacgt cacgccgacc tacgacagcg gcgagcgggc gggtgcgacc 600
atgcagcaca ttgcccgcca aggcgggctgc tgggtggtga gctgcggctc ggcgtttcag 660
gcgcgcgacg tcccggacgc gtttccgggg aagagcgagc ttttccgcga caacgacgag 720
tggtatcaacc cgggcgactc ggtcgtggtc gcgccgggcg gcaaggctcg cgccggggccg 780
ctgcacaaag aacgcgcgat cctgtacgcc gagatcgacc tcgagcgggt cggcgtggcg 840
cgccgcagcc tggacgtggt cggccattat gcgcggcccg acctcttcga cctgcacgtg 900
aacgcccgcc cgcaaagcgt gggtgaattg cgctga 936

```

<210> 68  
 <211> 311  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 68
Met Pro Arg Val Ala Val Val Gln Arg Pro Pro Val Phe Leu Asp Arg
 1          5          10          15
Ala Ala Thr Leu Glu Asn Ala Val Ala Ser Leu Ala Glu Ala Ala Ser
 20          25          30
Asn Gly Ala Arg Leu Ala Val Phe Pro Glu Ala Leu Val Pro Gly Tyr
 35          40          45
Pro Ala Trp Met Trp Arg Leu Arg Pro Gly Pro Asp Met Ala Leu Thr
 50          55          60
Glu Arg Ile His Ala Arg Leu Arg Ala Asn Ser Val Ser Leu Ala Ala
 65          70          75          80
Asp Glu Leu Ala Pro Leu Arg Glu Ala Ala Arg Arg His Glu Leu Thr
 85          90          95
Val Val Cys Gly Leu His Glu Arg Asp Glu Ala Leu Gly Gly Gly Thr
100          105          110
Leu Tyr Asn Thr Val Val Thr Ile Gly Ala Asp Gly Ala Val Leu Asn
115          120          125
Arg His Arg Lys Leu Met Pro Thr Asn Pro Glu Arg Met Val Trp Gly
130          135          140
Cys Gly Asp Ala Ser Gly Leu Arg Thr Val Pro Thr Gln Cys Gly Arg
145          150          155          160
Val Gly Ala Leu Ile Cys Trp Glu Ser Tyr Met Pro Leu Ala Arg Tyr
165          170          175
Ala Leu Tyr Ala Gln Gly Ile Asp Leu Tyr Val Thr Pro Thr Tyr Asp
180          185          190
Ser Gly Glu Arg Ala Val Ala Thr Met Gln His Ile Ala Arg Glu Gly
195          200          205
Gly Cys Trp Val Val Ser Cys Gly Ser Ala Phe Gln Ala Arg Asp Val
210          215          220
Pro Asp Ala Phe Pro Gly Lys Ser Glu Leu Phe Arg Asp Asn Asp Glu
225          230          235          240
Trp Ile Asn Pro Gly Asp Ser Val Val Val Ala Pro Gly Gly Lys Val
245          250          255
Val Ala Gly Pro Leu His Lys Glu Arg Ala Ile Leu Tyr Ala Glu Ile
260          265          270
Asp Leu Glu Arg Val Gly Val Ala Arg Arg Ser Leu Asp Val Val Gly
275          280          285
His Tyr Ala Arg Pro Asp Leu Phe Asp Leu His Val Asn Ala Arg Pro
290          295          300
Gln Ser Val Val Glu Leu Arg
305          310

```

<210> 69  
 <211> 939  
 <212> DNA



WO 03/097810

PCT/US03/15712

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 69

```

gtgaccgagt ttcggacggt gcgggtcgca gcggtgcagg cgacgccggt gaccctcgac      60
gccgatgcct cggtcgagaa ggcgatcggg ctgatcggcg aggcggtggc cgggtggagcg      120
cagctcgtcg tgctgcccga ggccttcgtg tcgctctacc cgtcgaacgc gtgggcgcga      180
gcgggccgccc gattcggcgg cttcgacgag ctctgggagc ggatgtgggc cagctcgctc      240
gacgtcccgg gcccgctggt cgaccggctg gtcgatgcgt gccgcaggca tgacgtggta      300
tgcgtgatcg gcgtgaacga gcgcgaaagc gaaaggccgg ggtcgcttta caacacgatg      360
ctgaccctcg gcccgtcggg cctcctgcac cggcaccgca agctcatgcc gacgcaccac      420
gagcggctgt tccatgggat cggcgacggt caagacctcg gcgttgtgga gaccgacgcg      480
ggacggatcg ggggactgat ctgctgggag aaccgaatgc cgctcgcgcg ctacgcggtc      540
taccagggtg gaccgcagat ctgggtcgcg ccgacggccg atgactccga cggctggctc      600
gcgagcatgc gccacatcgc gatcgagtcg ggcgcgttcg tcgtgtcggg gccgcagttc      660
atcccggcgt ccgcgttccc cgacgatttc cccgtcgagc taccgccggg caaggagggtg      720
ttcggccgcg gcggtgcggc gatcgtcgag ccgacctggg gcgaggtaat cgccggggccg      780
ctctacgatc gggaggggat cgtgttcgcc gactgtgacc tgcgacgcgg cttgcatgcc      840
aagcgctggt tcgactccgt cggccattac agccgcgcgg aggtgctcga tggcggcgctc      900
gagcgcgtcc cggcgccggt ggacggcgaa tcgccgtga      939

```

&lt;210&gt; 70

&lt;211&gt; 312

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 70

```

Val Thr Glu Phe Arg Thr Val Arg Val Ala Ala Val Gln Ala Thr Pro
 1          5          10          15
Val Thr Leu Asp Ala Asp Ala Ser Val Glu Lys Ala Ile Gly Leu Ile
          20          25          30
Gly Glu Ala Val Ala Gly Gly Ala Gln Leu Val Val Leu Pro Glu Ala
          35          40          45
Phe Val Ser Leu Tyr Pro Ser Asn Ala Trp Ala Arg Ala Ala Ala Gly
          50          55          60
Phe Gly Gly Phe Asp Glu Leu Trp Glu Arg Met Trp Ala Ser Ser Leu
65          70          75          80
Asp Val Pro Gly Pro Leu Val Asp Arg Leu Val Asp Ala Cys Arg Arg
          85          90          95
His Asp Val Val Cys Val Ile Gly Val Asn Glu Arg Glu Ser Glu Arg
          100          105          110
Pro Gly Ser Leu Tyr Asn Thr Met Leu Thr Leu Gly Pro Ser Gly Leu
          115          120          125
Leu His Arg His Arg Lys Leu Met Pro Thr His His Glu Arg Leu Phe
          130          135          140
His Gly Ile Gly Asp Gly Gln Asp Leu Gly Val Val Glu Thr Asp Ala
145          150          155          160
Gly Arg Ile Gly Gly Leu Ile Cys Trp Glu Asn Arg Met Pro Leu Ala
          165          170          175
Arg Tyr Ala Val Tyr Gln Gly Gly Pro Gln Ile Trp Val Ala Pro Thr
          180          185          190
Ala Asp Asp Ser Asp Gly Trp Leu Ala Ser Met Arg His Ile Ala Ile
          195          200          205
Glu Ser Gly Ala Phe Val Val Ser Val Pro Gln Phe Ile Pro Ala Ser
210          215          220
Ala Phe Pro Asp Asp Phe Pro Val Glu Leu Pro Pro Gly Lys Glu Val

```







|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |  |
| Ala | Glu | Ala | Ser | Gly | Ala | Ser | Met | Tyr | Asn | Thr | Leu | Leu | Tyr | Ile | Asp |  |  |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |  |
| Ala | Gln | Gly | Glu | Ile | Leu | Gly | Lys | His | Arg | Lys | Leu | Val | Pro | Thr | Gly |  |  |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |  |
| Gly | Glu | Arg | Leu | Val | Trp | Ala | Gln | Gly | Asp | Gly | Ser | Thr | Leu | Gln | Val |  |  |
| 145 |     |     | 150 |     |     |     |     | 155 |     |     |     |     |     | 160 |     |  |  |
| Tyr | Asp | Thr | Pro | Leu | Gly | Lys | Leu | Gly | Gly | Leu | Ile | Cys | Trp | Glu | Asn |  |  |
|     |     |     | 165 |     |     |     | 170 |     |     |     |     |     |     | 175 |     |  |  |
| Tyr | Met | Pro | Leu | Ala | Arg | Tyr | Thr | Met | Tyr | Ala | Trp | Gly | Thr | Gln | Ile |  |  |
|     |     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |     |  |  |
| Tyr | Val | Ala | Ala | Thr | Trp | Asp | Arg | Gly | Gln | Pro | Trp | Leu | Ser | Thr | Leu |  |  |
|     | 195 |     |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |  |  |
| Arg | His | Ile | Ala | Lys | Glu | Gly | Arg | Val | Tyr | Val | Ile | Gly | Cys | Cys | Ile |  |  |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |  |
| Ala | Met | Arg | Lys | Asp | Asp | Ile | Pro | Asp | His | Tyr | Thr | Met | Lys | Gln | Lys |  |  |
| 225 |     |     | 230 |     |     |     |     | 235 |     |     |     |     |     | 240 |     |  |  |
| Phe | Tyr | Ser | Asp | Ala | Asp | Glu | Trp | Ile | Asn | Ile | Gly | Asp | Ser | Ala | Ile |  |  |
|     |     |     | 245 |     |     |     | 250 |     |     |     |     |     |     | 255 |     |  |  |
| Val | Asn | Pro | Glu | Gly | Gln | Phe | Ile | Ala | Gly | Pro | Val | Arg | Lys | Gln | Glu |  |  |
|     | 260 |     |     |     |     | 265 |     |     |     |     |     | 270 |     |     |     |  |  |
| Glu | Ile | Leu | Tyr | Ala | Glu | Ile | Asp | Pro | Arg | Met | Val | Gln | Gly | Pro | Lys |  |  |
|     | 275 |     |     |     |     | 280 |     |     |     |     |     | 285 |     |     |     |  |  |
| Trp | Met | Leu | Asp | Val | Ala | Gly | His | Tyr | Ala | Arg | Pro | Asp | Val | Phe | Glu |  |  |
|     | 290 |     |     |     | 295 |     |     |     |     |     | 300 |     |     |     |     |  |  |
| Leu | Ile | Val | His | Thr | Asp | Ile | Arg | Arg | Met | Ile | Lys | Ser | Glu | Lys | Asn |  |  |
| 305 |     |     | 310 |     |     |     |     | 315 |     |     |     |     |     | 320 |     |  |  |
| Ser |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |

&lt;210&gt; 73

&lt;211&gt; 1035

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 73

|            |            |             |             |             |             |      |
|------------|------------|-------------|-------------|-------------|-------------|------|
| atgacagcaa | tagactcaac | gttttaaagtc | gccgccgttc  | aggctgcgcc  | ggtcttcctc  | 60   |
| aatcgcgacg | caaccgtgga | gaaggcgtgc  | cggctgatca  | agtccgcggc  | agagggaggc  | 120  |
| gcgcgtctga | tcgttttccc | ggaagcgttc  | ataccggcct  | acccggaactg | ggtgtggacg  | 180  |
| gtccctgccg | gtgagcaagg | cctgctcaac  | gacctctacg  | gccaaactcgt | cgaccagtcc  | 240  |
| gtgacgattc | ccagcgacat | caccaccgag  | ttatgtaacg  | cggcacgggc  | agcaaacgcc  | 300  |
| tatgtcgtga | ttggtgtcaa | cgagcgcaac  | gcggaggcaa  | gcaatggaag  | cctctacaac  | 360  |
| tcgctcctct | acatcgacgc | aaacggcaaa  | attctcggta  | agcacgcgaa  | gctcgttccc  | 420  |
| acaggcggag | aacggctcgt | gtgggcgcag  | ggcgatggca  | gcacgctcga  | agcctacgac  | 480  |
| acggagctgg | gcaaactcgg | cggtctcatt  | tgctgggaga  | actatatgcc  | gctggcacgc  | 540  |
| tacgcgatgt | acgcatgggg | agtgcagctc  | tatgtcgccg  | cgacctggga  | ccgtggcggc  | 600  |
| ccctggactg | ccacgctgcg | tcatgtcgcc  | aaggaaggtc  | agatgtacgt  | catcgggtgc  | 660  |
| tgccaggccc | tgcacaagga | tgacctgccg  | gagctagacg  | ggctgaagga  | gaagtactac  | 720  |
| gccaacgcac | gagagtggat | caatgttggc  | gacagcgcta  | ttgtcggccc  | ggacggacaa  | 780  |
| ttccttgtcg | agcccgtccg | aatgcgggaa  | gacatcctct  | acgccgaggt  | ggacactcgc  | 840  |
| aacttccgcg | gcccgaagtg | gatgttcgac  | gcggctggac  | actacgcgcg  | tcccgcacatt | 900  |
| ttccaactca | cagtgaaccg | cgagcagcgg  | ccgatgggtcc | gcgtcgtcgg  | tgacagcagt  | 960  |
| gaccagaagg | agcggccgct | cccggacgac  | ggacggctct  | ggtacgccta  | cagcaccaat  | 1020 |
| cagcaccacg | actga      |             |             |             |             | 1035 |

&lt;210&gt; 74

&lt;211&gt; 344

&lt;212&gt; PRT



WO 03/097810

PCT/US03/15712

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 74

```

Met Thr Ala Ile Asp Ser Thr Phe Lys Val Ala Ala Val Gln Ala Ala
 1          5          10          15
Pro Val Phe Leu Asn Arg Asp Ala Thr Val Glu Lys Ala Cys Arg Leu
          20          25          30
Ile Lys Ser Ala Ala Glu Gly Gly Ala Arg Leu Ile Val Phe Pro Glu
          35          40          45
Ala Phe Ile Pro Ala Tyr Pro Asp Trp Val Trp Thr Val Pro Ala Gly
          50          55          60
Glu Gln Gly Leu Leu Asn Asp Leu Tyr Gly Gln Leu Val Asp Gln Ser
65          70          75          80
Val Thr Ile Pro Ser Asp Ile Thr Thr Glu Leu Cys Asn Ala Ala Arg
          85          90          95
Ala Ala Asn Ala Tyr Val Val Ile Gly Val Asn Glu Arg Asn Ala Glu
          100          105          110
Ala Ser Asn Gly Ser Leu Tyr Asn Ser Leu Leu Tyr Ile Asp Ala Asn
          115          120          125
Gly Lys Ile Leu Gly Lys His Arg Lys Leu Val Pro Thr Gly Gly Glu
          130          135          140
Arg Leu Val Trp Ala Gln Gly Asp Gly Ser Thr Leu Glu Ala Tyr Asp
145          150          155          160
Thr Glu Leu Gly Lys Leu Gly Gly Leu Ile Cys Trp Glu Asn Tyr Met
          165          170          175
Pro Leu Ala Arg Tyr Ala Met Tyr Ala Trp Gly Val Gln Leu Tyr Val
          180          185          190
Ala Ala Thr Trp Asp Arg Gly Gly Pro Trp Thr Ala Thr Leu Arg His
          195          200          205
Val Ala Lys Glu Gly Gln Met Tyr Val Ile Gly Cys Cys Gln Ala Leu
          210          215          220
His Lys Asp Asp Leu Pro Glu Leu Asp Gly Leu Lys Glu Lys Tyr Tyr
225          230          235          240
Ala Asn Ala Arg Glu Trp Ile Asn Val Gly Asp Ser Ala Ile Val Gly
          245          250          255
Pro Asp Gly Gln Phe Leu Val Glu Pro Val Arg Met Arg Glu Asp Ile
          260          265          270
Leu Tyr Ala Glu Val Asp Thr Arg Asn Phe Arg Gly Pro Lys Trp Met
          275          280          285
Phe Asp Ala Ala Gly His Tyr Ala Arg Pro Asp Ile Phe Gln Leu Thr
          290          295          300
Val Asn Arg Glu Gln Arg Pro Met Val Arg Val Val Gly Asp Ser Ser
305          310          315          320
Asp Gln Lys Glu Arg Pro Leu Pro Asp Asp Gly Arg Leu Trp Tyr Ala
          325          330          335
Tyr Ser Thr Asn Gln His His Asp
          340

```

&lt;210&gt; 75

&lt;211&gt; 1125

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 75

atgagcacca ttgttaaagc cgctgogggtt caaatcagcc cagtcctcta cagccgag

60



```

gggacagtcg caaaagttgt gcggaagatc cacgaacttg gccaaaaggg ggtgcggttc 120
gccacgttcc cggagaccgt ggttccttac tatccatatt tttccgccgt ccagaccccc 180
attcaactat tgtccggaac cgagtacctg aagttgctcg accaaggcgt gaccgtgccg 240
tccacgacta ccgacgcaat cggggaggct gcccggaacg ccggcatggt tgtatctatc 300
ggcgtgaatg agcgtgacgg cgggaccctg tacaacgcgc agttgctctt cgatgcggat 360
gggaccttga ttcagcgtcg ccgcaagatc actcctacgc attacgagcg catgatctgg 420
ggccagggag atggttcggg tttgcgggcc gtcaagagcc aggttggtcg tattggccaa 480
cttgcatgct ttgagcacia caaccactg gcgcgttacg cgatgatggc cgatggcgag 540
caaatccatt cggccatgta tccaggttcc gcgttcggcg aggggttcgc ggaaaagatg 600
gaaatcaata tccgccagca tgcgttgagg tccgggtgct tcgttggtgaa tgcaacggcc 660
tggcttgacg ccagccagca ggcacaaatc atgaatgaca cgggttgcca aatcgggtccg 720
atctcgggcg gttgctttac cacgatcgta acaccgacg gcacgtttct gggcgaacct 780
ctccggtcgg gtgagggcga ggtcatcgcc gatctcgatt tcaagctgat cgacaaacgc 840
aagatgttga tggactcgcg cggccactac agtcgcccgg aattgctcag tctgctgatc 900
gaccgcaccc ccaccgcgca cattcatgag cgaggtgcgc cgcagacgtc aggcgctgtg 960
caagaggcga cgaaagtggg ttcacacgcg ccgctcctgc gtgacggaca atgggatcag 1020
ctcaatgcgg gagcggggccg acatacaggg aatggagaag cacagataga aatcatggcc 1080
gcggcccact cgggcacccg tggaattgaa gcgaaggagg cctaa 1125

```

<210> 76  
 <211> 374  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 76

```

Met Ser Thr Ile Val Lys Ala Ala Ala Val Gln Ile Ser Pro Val Leu
  1          5          10          15
Tyr Ser Arg Glu Gly Thr Val Ala Lys Val Val Arg Lys Ile His Glu
      20      25      30
Leu Gly Gln Lys Gly Val Arg Phe Ala Thr Phe Pro Glu Thr Val Val
      35      40      45
Pro Tyr Tyr Pro Tyr Phe Ser Ala Val Gln Thr Pro Ile Gln Leu Leu
      50      55      60
Ser Gly Thr Glu Tyr Leu Lys Leu Leu Asp Gln Gly Val Thr Val Pro
65      70      75      80
Ser Thr Thr Thr Asp Ala Ile Gly Glu Ala Ala Arg Asn Ala Gly Met
      85      90      95
Val Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Leu Tyr Asn
      100      105      110
Ala Gln Leu Leu Phe Asp Ala Asp Gly Thr Leu Ile Gln Arg Arg Arg
      115      120      125
Lys Ile Thr Pro Thr His Tyr Glu Arg Met Ile Trp Gly Gln Gly Asp
      130      135      140
Gly Ser Gly Leu Arg Ala Val Lys Ser Gln Val Gly Arg Ile Gly Gln
145      150      155      160
Leu Ala Cys Phe Glu His Asn Asn Pro Leu Ala Arg Tyr Ala Met Met
      165      170      175
Ala Asp Gly Glu Gln Ile His Ser Ala Met Tyr Pro Gly Ser Ala Phe
      180      185      190
Gly Glu Gly Phe Ala Glu Lys Met Glu Ile Asn Ile Arg Gln His Ala
      195      200      205
Leu Glu Ser Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala
      210      215      220
Ser Gln Gln Ala Gln Ile Met Asn Asp Thr Gly Cys Gln Ile Gly Pro
225      230      235      240
Ile Ser Gly Gly Cys Phe Thr Thr Ile Val Thr Pro Asp Gly Thr Phe
      245      250      255
Leu Gly Glu Pro Leu Arg Ser Gly Glu Gly Glu Val Ile Ala Asp Leu

```



```

                260                265                270
Asp Phe Lys Leu Ile Asp Lys Arg Lys Met Leu Met Asp Ser Arg Gly
                275                280                285
His Tyr Ser Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Thr Pro
                290                295                300
Thr Ala His Ile His Glu Arg Gly Ala Pro Gln Thr Ser Gly Ala Val
305                310                315                320
Gln Glu Ala Thr Lys Val Gly Ser His Ala Pro Leu Leu Arg Asp Gly
                325                330                335
Gln Trp Asp Gln Leu Asn Ala Gly Ala Gly Arg His Thr Gly Asn Gly
                340                345                350
Glu Ala Gln Ile Glu Ile Met Ala Ala Ala His Ser Gly Thr Arg Gly
                355                360                365
Ile Glu Ala Lys Gly Ala
                370

```

&lt;210&gt; 77

&lt;211&gt; 1056

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 77

```

atgccaaaccc ccagcgatca tttcaaaatc gccgctgttc aggcctcgcc cgtgtttctg      60
gaccgggagg ccactgtgga aaaggcctgc cgttgatcg ccgaagccgc aaagcagggc      120
gtccgcctca tcgtctttcc ggaatcggtc atcccgacct acccgactg ggtatgggccc      180
gttcccccgga gaagggaag aatcctgaac cagctgtatt ctgaattcct ggccaatgcc      240
gtcgatgttc ccggcgcggc gaccgaacaa cttgcccagg ctgcacgaat ggccggcgcc      300
tatgtgatta tgggcgtcac cgaaagagac acctcgcca gcggggccag cctctacaac      360
accctgctct acttcagccc cgaaggcatc ctaatgggca aacaccggaa gctggttccc      420
acggggggcg aacggctggt ctgggcctac ggagacggca gcacgctgga ggtctacgac      480
actccgctgg gaaagatcgg cgggctgatc tgctgggaga actacatgcc cctggcccgg      540
tacacgatgt acgcctgggg caccagatt tacatcgccg ccacctggga ccgcggggaa      600
ccgtggctct ccaccctgcg gcatatcgca aaggaaggaa gggctctacgt catcgggtgc      660
tgcacgccc tgcgccaggg ggatatccc gaccggttcg agtacaaggg aaaattttat      720
tccgggtccc gggagtggat caatgagggc gacagcgcca tcgtgaaccc ggacggggaa      780
ttcatcgccg ggccgggtgc gacgaaggag gagatcctgt atgccgagat agacccccgg      840
cagatgcggg gccccaagt gatgctcgat gtggccggtc attacgccc gccggatatc      900
ttcgagctca tcgtccaccg gaatccccac ccgatgatca aaatcgccga agacaggggc      960
acgggggatcg cctcaagttt gattcgcccc cgccctaacc ttcccccatc aagggggagg     1020
aaatcggcaa gaagcaaacg caagcccaaa aaatga                                1056

```

&lt;210&gt; 78

&lt;211&gt; 351

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 78

```

Met Pro Thr Pro Ser Asp His Phe Lys Ile Ala Ala Val Gln Ala Ser
 1                5                10                15
Pro Val Phe Leu Asp Arg Glu Ala Thr Val Glu Lys Ala Cys Arg Leu
                20                25                30
Ile Ala Glu Ala Ala Lys Gln Gly Val Arg Leu Ile Val Phe Pro Glu
                35                40                45
Ser Phe Ile Pro Thr Tyr Pro Asp Trp Val Trp Ala Val Pro Pro Gly
                50                55                60

```



Arg Glu Arg Ile Leu Asn Gln Leu Tyr Ser Glu Phe Leu Ala Asn Ala  
 65 70 75 80  
 Val Asp Val Pro Gly Ala Ala Thr Glu Gln Leu Ala Gln Ala Ala Arg  
 85 90 95  
 Met Ala Gly Ala Tyr Val Ile Met Gly Val Thr Glu Arg Asp Thr Ser  
 100 105 110  
 Ala Ser Gly Ala Ser Leu Tyr Asn Thr Leu Leu Tyr Phe Ser Pro Glu  
 115 120 125  
 Gly Ile Leu Met Gly Lys His Arg Lys Leu Val Pro Thr Gly Gly Glu  
 130 135 140  
 Arg Leu Val Trp Ala Tyr Gly Asp Gly Ser Thr Leu Glu Val Tyr Asp  
 145 150 155 160  
 Thr Pro Leu Gly Lys Ile Gly Gly Leu Ile Cys Trp Glu Asn Tyr Met  
 165 170 175  
 Pro Leu Ala Arg Tyr Thr Met Tyr Ala Trp Gly Thr Gln Ile Tyr Ile  
 180 185 190  
 Ala Ala Thr Trp Asp Arg Gly Glu Pro Trp Leu Ser Thr Leu Arg His  
 195 200 205  
 Ile Ala Lys Glu Gly Arg Val Tyr Val Ile Gly Cys Cys Ile Ala Leu  
 210 215 220  
 Arg Gln Gly Asp Ile Pro Asp Arg Phe Glu Tyr Lys Gly Lys Phe Tyr  
 225 230 235 240  
 Ser Gly Ser Arg Glu Trp Ile Asn Glu Gly Asp Ser Ala Ile Val Asn  
 245 250 255  
 Pro Asp Gly Glu Phe Ile Ala Gly Pro Val Arg Thr Lys Glu Glu Ile  
 260 265 270  
 Leu Tyr Ala Glu Ile Asp Pro Arg Gln Met Arg Gly Pro Lys Trp Met  
 275 280 285  
 Leu Asp Val Ala Gly His Tyr Ala Arg Pro Asp Ile Phe Glu Leu Ile  
 290 295 300  
 Val His Arg Asn Pro His Pro Met Ile Lys Ile Ala Glu Asp Arg Gly  
 305 310 315 320  
 Thr Gly Ile Ala Ser Ser Leu Ile Arg Pro Arg Pro Asn Leu Pro Pro  
 325 330 335  
 Ser Arg Gly Arg Lys Ser Ala Arg Ser Lys Arg Lys Pro Lys Lys  
 340 345 350

<210> 79  
 <211> 990  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 79  
 atgacgaaga aaagcggccg cgattcgttt cgggtcgctg cggtccaggc ctctgctcgtc 60  
 tacctggatc gggaacggag catcgagaaa gcgtgccggc tgatcgacga cgcgggacga 120  
 aacgacgccg acctcgctgt gttccccgaa gccttcgtgc ccggataccc actgtgggtg 180  
 tggctcgttc cgccggggcg caccgcagac ttgcgctccg cttatgacga gctccacgcc 240  
 aacgcgatca gcattccgga cgactccacc gatcggctgt gcgccgccgc aaaagacgcc 300  
 ggcgtcgccg tcgcgatcgg cgtcaacgaa cgcaacaccg aagcgagcgg catgagcctg 360  
 ttcaacacgc tgctctatat cggagcggac ggccggattc tcggaaaaca ccggaagctg 420  
 gtaccgaccg gcggcgaacg gctcgtcttg gcattctggc acggcagcga cctcgaggtc 480  
 tactcgtgc cgctcggtcg cgtaagcggg ctgatctgct gggagcacta catgccgctc 540  
 gcccggtatg cgctcgccgc gtggggcgaa cagggtcacg tcgctccaac ctgggatcgt 600  
 ggcgagccgt ggctgtccac gctaaggcac atcgcgaaag aaggccgcgt tctcgtcgtc 660  
 ggctgctgtc aagccgtgcg caaggacgac atccctgaca cgctcgcgtt caagtccaaa 720  
 tacctcgacg acgtggacgg ctggatcaac ccagggtggc gcgtcatcat caatcctgac 780  
 ggcaaggctc tcgcgggacc ggcgatggaa accgaaactg tactgtacgc ggaccttcgc 840  
 accgagcagc tcgtcggacc gcgctggcag ctcgacgtcg gcggacatta cgctcgtccg 900



WO 03/097810

PCT/US03/15712

gacgtcttcg agctcgtcgt ccacggcat ccgaagccgt tgattcggac agcgaccggt 960  
 gtcaggcgcc gcaagcgtgc acgtcgctaa 990

<210> 80  
 <211> 329  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 80  
 Met Thr Lys Lys Ser Gly Arg Asp Ser Phe Arg Val Ala Ala Val Gln  
   1                  5                  10                  15  
 Ala Ser Ser Val Tyr Leu Asp Arg Glu Arg Ser Ile Glu Lys Ala Cys  
                   20                  25                  30  
 Arg Leu Ile Asp Asp Ala Gly Arg Asn Asp Ala Asp Leu Val Val Phe  
                   35                  40                  45  
 Pro Glu Ala Phe Val Pro Gly Tyr Pro Leu Trp Val Trp Leu Val Pro  
                   50                  55                  60  
 Pro Gly Arg Thr Ala Asp Leu Arg Ser Ala Tyr Ala Thr Leu His Ala  
   65                  70                  75                  80  
 Asn Ala Ile Ser Ile Pro Asp Asp Ser Thr Asp Arg Leu Cys Ala Ala  
                   85                  90                  95  
 Ala Lys Asp Ala Gly Val Ala Val Ala Ile Gly Val Asn Glu Arg Asn  
                   100                  105                  110  
 Thr Glu Ala Ser Gly Met Ser Leu Phe Asn Thr Leu Leu Tyr Ile Gly  
                   115                  120                  125  
 Ala Asp Gly Arg Ile Leu Gly Lys His Arg Lys Leu Val Pro Thr Gly  
                   130                  135                  140  
 Gly Glu Arg Leu Val Trp Ala Ser Gly Asp Gly Ser Asp Leu Glu Val  
   145                  150                  155                  160  
 Tyr Ser Leu Pro Phe Gly Arg Val Ser Gly Leu Ile Cys Trp Glu His  
                   165                  170                  175  
 Tyr Met Pro Leu Ala Arg Tyr Ala Leu Ala Ala Trp Gly Glu Gln Val  
                   180                  185                  190  
 His Val Ala Pro Thr Trp Asp Arg Gly Glu Pro Trp Leu Ser Thr Leu  
                   195                  200                  205  
 Arg His Ile Ala Lys Glu Gly Arg Val Leu Val Val Gly Cys Cys Gln  
                   210                  215                  220  
 Ala Val Arg Lys Asp Asp Ile Pro Asp Thr Leu Ala Phe Lys Ser Lys  
   225                  230                  235                  240  
 Tyr Leu Ala Asp Val Asp Gly Trp Ile Asn Pro Gly Gly Ser Val Ile  
                   245                  250                  255  
 Ile Asn Pro Asp Gly Lys Val Val Ala Gly Pro Ala Met Glu Thr Glu  
                   260                  265                  270  
 Thr Val Leu Tyr Ala Asp Leu Arg Thr Glu Gln Leu Val Gly Pro Arg  
                   275                  280                  285  
 Trp Gln Leu Asp Val Gly Gly His Tyr Ala Arg Pro Asp Val Phe Glu  
                   290                  295                  300  
 Leu Val Val His Arg His Pro Lys Pro Leu Ile Arg Thr Ala Thr Gly  
   305                  310                  315                  320  
 Val Arg Arg Arg Lys Arg Ala Arg Arg  
                   325

<210> 81  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>



<223> Obtained from an environmental sample

<400> 81  
 atgaaagtcg tcaaagccgc cgctgtccag ttcagcccgg tgctctatag ccgcgaagcg 60  
 accgtcgcca aggtcggtccg gaaaatccac gagctcggtc agaaaggcgt gcagttcgcc 120  
 acctttcctg aaacgggtcgt gccttattac cttacttcg cggccgtcca gacgggcatc 180  
 gagctcttgt cgggcaccga acatctgcgc ctgctcgaac aggcctgac tgtgccctcc 240  
 gctgcgaccg atgcaatcgg cgaagccgcg cgacaggccg gcatggtcgt gtccatcggc 300  
 gtcaatgagc gtgacggcgg cacgctttac aacacgcaac tgctcttcga tgccgacggt 360  
 acgctgatcc agcggccgccc caagatcacg ccgacccatt tcgaacgcat gatctggggg 420  
 cagggagatg gctcgggctt gcgtgcagtc gacagcgagc tcggccgcat cggccagctc 480  
 gcatgcttcg agcacaacaa cccgcttgca cgttacgcaa tgatcgccga cggcgagcag 540  
 atccattcag cgatgtaccc tggctcggcc tttggcgagg gcttcgcccga gcgtatggag 600  
 atcaacatcc gccagcatgc gctcgagtc gccgctttcg tcgtcaacgc aacggcgtgg 660  
 cttgacgccg accagcaggc gcaaatcatg aaggacaccg gttgtggaat cgggtccgatc 720  
 tcgggcggct gcttcaccac gatcgtttct cctgacggta tgctgatggc cgatccgctt 780  
 cgctcgggcg aaggcgaagt gattgtcgat ctcgacttca cgcagatcga ccgccgcaag 840  
 atgctgatgg actcggccgg ccactacaac cgccctgaac tgctgagtct gatgatcgac 900  
 cgtacgccgg ctgcgcatgt tcacgaacgc gcttcgcgcc cgatgaccgt cgacgaccag 960  
 agttccggcg atctgcgcac ccagggttgca tga 993

<210> 82  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 82  
 Met Lys Val Val Lys Ala Ala Ala Val Gln Phe Ser Pro Val Leu Tyr  
 1 5 10 15  
 Ser Arg Glu Ala Thr Val Ala Lys Val Val Arg Lys Ile His Glu Leu  
 20 25 30  
 Gly Gln Lys Gly Val Gln Phe Ala Thr Phe Pro Glu Thr Val Val Pro  
 35 40 45  
 Tyr Tyr Pro Tyr Phe Ala Ala Val Gln Thr Gly Ile Glu Leu Leu Ser  
 50 55 60  
 Gly Thr Glu His Leu Arg Leu Leu Glu Gln Ala Val Thr Val Pro Ser  
 65 70 75 80  
 Ala Ala Thr Asp Ala Ile Gly Glu Ala Ala Arg Gln Ala Gly Met Val  
 85 90 95  
 Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Leu Tyr Asn Thr  
 100 105 110  
 Gln Leu Leu Phe Asp Ala Asp Gly Thr Leu Ile Gln Arg Arg Arg Lys  
 115 120 125  
 Ile Thr Pro Thr His Phe Glu Arg Met Ile Trp Gly Gln Gly Asp Gly  
 130 135 140  
 Ser Gly Leu Arg Ala Val Asp Ser Ala Val Gly Arg Ile Gly Gln Leu  
 145 150 155 160  
 Ala Cys Phe Glu His Asn Asn Pro Leu Ala Arg Tyr Ala Met Ile Ala  
 165 170 175  
 Asp Gly Glu Gln Ile His Ser Ala Met Tyr Pro Gly Ser Ala Phe Gly  
 180 185 190  
 Glu Gly Phe Ala Gln Arg Met Glu Ile Asn Ile Arg Gln His Ala Leu  
 195 200 205  
 Glu Ser Ala Ala Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala Asp  
 210 215 220  
 Gln Gln Ala Gln Ile Met Lys Asp Thr Gly Cys Gly Ile Gly Pro Ile  
 225 230 235 240  
 Ser Gly Gly Cys Phe Thr Thr Ile Val Ser Pro Asp Gly Met Leu Met



|          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <400> 84 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Met      | Met | Ser | Ser | Ala | Arg | Val | Ile | Lys | Leu | Ala | Ala | Ala | Gln | Leu | Ser |
| 1        |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Pro      | Val | Leu | Pro | Gly | Glu | Ser | Thr | Asn | Ser | Arg | Asp | Gly | Thr | Ile | Ala |
|          |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Lys      | Val | Val | Ala | Ala | Ile | Ala | Glu | Ala | Ala | Arg | Ala | Gly | Ala | Gln | Leu |
|          |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Ile      | Val | Phe | Pro | Glu | Thr | Val | Val | Pro | Tyr | Tyr | Pro | Tyr | Phe | Ser | Phe |
|          | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Ile      | Thr | Pro | Ala | Val | Thr | Met | Gly | Ala | Glu | His | Leu | Arg | Leu | Tyr | Asp |
| 65       |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Gln      | Ser | Val | Val | Val | Pro | Ser | Ala | Ala | Thr | Asp | Thr | Val | Ala | Ala | Ala |
|          |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |



Ala Lys Lys His Ser Met Val Val Val Leu Gly Ile Asn Glu Arg Asp  
 100 105 110  
 His Gly Thr Leu Tyr Asn Ala Gln Leu Ile Phe Asp Ala Ser Gly Glu  
 115 120 125  
 Leu Leu Leu Lys Arg Arg Lys Ile Thr Pro Thr Tyr His Glu Arg Met  
 130 135 140  
 Val Trp Gly Gln Gly Asp Gly Ser Gly Leu Lys Thr Val Asp Thr Ala  
 145 150 155 160  
 Ile Gly Arg Val Gly Ala Leu Ala Cys Trp Glu His Tyr Asn Pro Leu  
 165 170 175  
 Ala Arg Tyr Ser Leu Met Ala Gln His Glu Glu Ile His Cys Ser Gln  
 180 185 190  
 Phe Pro Gly Ser Leu Val Gly Pro Ile Phe Ala Glu Gln Met Glu Val  
 195 200 205  
 Thr Met Arg His His Ala Leu Glu Ser Gly Cys Phe Val Val Asn Ala  
 210 215 220  
 Thr Ala Trp Leu Ser Glu Ala Gln Ile Gln Ser Ile Ser Ser Asp Pro  
 225 230 235 240  
 Ala Met Gln Lys Ala Leu Arg Gly Gly Cys Tyr Thr Ala Ile Ile Ser  
 245 250 255  
 Pro Glu Gly Lys His Leu Cys Glu Pro Leu Arg Glu Gly Glu Gly Leu  
 260 265 270  
 Ile Phe Ala Glu Ala Asp Met Ala Leu Ile Thr Lys Arg Lys Arg Met  
 275 280 285  
 Met Asp Ser Val Gly His Tyr Ala Arg Pro Glu Leu Leu Ser Leu Leu  
 290 295 300  
 Ile Asp His Arg Ala Thr Thr Pro Leu His Ser Val Thr Ala Ser Asp  
 305 310 315 320  
 Ala Ala Ala Val Lys Asn Thr Arg Ser Ser Ala His Glu Ser Ala Asp  
 325 330 335  
 Ser Glu Thr Ile Arg Glu Ser Val Asn Asn Gly Thr Pro Ile Ala Arg  
 340 345 350  
 Leu Ala Pro Ser  
 355

<210> 85  
 <211> 1014  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 85  
 atgggtctggt ttcattcagaa atacaagggtt gcggtggttc aggcggcgcc ggtctttctc 60  
 gacctcgatg cgacgggtgga caagacgatc gccctgatcg agcaggccgc agcacagggc 120  
 gcgaagctga tcgcgtttcc cgagaccttc attcccggat atccgtggca gatctggctt 180  
 ggggcgcccg cctgggcgat cggccgtggc ttcgtgcagc gctatttcga taactcgttg 240  
 tcatttgaca gcccgcaggc cgaaaaaatt cgcaaggccg tcaagcgcgc caagctgacc 300  
 gcggtgatcg gcgtctccga acgcgacggc ggcagcctct atatcggccca atggctgatc 360  
 ggtcccgcag gcgagaccat tgcgaagcgc cgcaagctgc ggccgaccca tgccgaacgc 420  
 accgtgttcg gcgagggcga cggcagcgac ctgcgccgtc atgatcgcg cgacgtggga 480  
 cggctcggtg caatgtgctg ctgggagcat ctgcagccgc tgtcgaaata cgcgatgtac 540  
 gcccagaacg agcaggttca cgtcggcgcc tggccgagct tctcattgta cgacccattc 600  
 gcccatgcgc ttggctggga agtaaacaac gcggcgagca aggtttatgc tgtcgagggc 660  
 tcatgtttct tcctcggccc gtgcgcggtg gtctcgcagg ccatgatcga cgagctctgc 720  
 gattcccccg aaaagcacgc cttcctgcac gctggcgggc gccacgcggt aatctatggg 780  
 ccggacggga gttcgcttgc cgacaaactt ccacccgatc aggagggcat tctgtatgcc 840  
 gatatcgatc tcggcatgat cggcgtggca aagaacgccg ccgacccgcg aggacactat 900  
 tccaggccgg acgtcacgcg gctgctgctc aacacttccc gcgccaatcg cgtcgagcat 960  
 ttttcattgc cgatcgatgc cgaggatcat agcgaaatca gacttcaggc ctga 1014



WO 03/097810

PCT/US03/15712

<210> 86  
 <211> 337  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 86  
 Met Gly Leu Val His Gln Lys Tyr Lys Val Ala Val Val Gln Ala Ala  
 1 5 10 15  
 Pro Val Phe Leu Asp Leu Asp Ala Thr Val Asp Lys Thr Ile Ala Leu  
 20 25 30  
 Ile Glu Gln Ala Ala Ala Gln Gly Ala Lys Leu Ile Ala Phe Pro Glu  
 35 40 45  
 Thr Phe Ile Pro Gly Tyr Pro Trp Gln Ile Trp Leu Gly Ala Pro Ala  
 50 55 60  
 Trp Ala Ile Gly Arg Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu  
 65 70 75 80  
 Ser Phe Asp Ser Pro Gln Ala Glu Lys Ile Arg Lys Ala Val Lys Arg  
 85 90 95  
 Ala Lys Leu Thr Ala Val Ile Gly Val Ser Glu Arg Asp Gly Gly Ser  
 100 105 110  
 Leu Tyr Ile Gly Gln Trp Leu Ile Gly Pro Asp Gly Glu Thr Ile Ala  
 115 120 125  
 Lys Arg Arg Lys Leu Arg Pro Thr His Ala Glu Arg Thr Val Phe Gly  
 130 135 140  
 Glu Gly Asp Gly Ser Asp Leu Ala Val His Asp Arg Ala Asp Val Gly  
 145 150 155 160  
 Arg Leu Gly Ala Met Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys  
 165 170 175  
 Tyr Ala Met Tyr Ala Gln Asn Glu Gln Val His Val Gly Ala Trp Pro  
 180 185 190  
 Ser Phe Ser Leu Tyr Asp Pro Phe Ala His Ala Leu Gly Trp Glu Val  
 195 200 205  
 Asn Asn Ala Ala Ser Lys Val Tyr Ala Val Glu Gly Ser Cys Phe Phe  
 210 215 220  
 Leu Gly Pro Cys Ala Val Val Ser Gln Ala Met Ile Asp Glu Leu Cys  
 225 230 235 240  
 Asp Ser Pro Glu Lys His Ala Phe Leu His Ala Gly Gly Gly His Ala  
 245 250 255  
 Val Ile Tyr Gly Pro Asp Gly Ser Ser Leu Ala Asp Lys Leu Pro Pro  
 260 265 270  
 Asp Gln Glu Gly Ile Leu Tyr Ala Asp Ile Asp Leu Gly Met Ile Gly  
 275 280 285  
 Val Ala Lys Asn Ala Ala Asp Pro Ala Gly His Tyr Ser Arg Pro Asp  
 290 295 300  
 Val Thr Arg Leu Leu Leu Asn Thr Ser Arg Ala Asn Arg Val Glu His  
 305 310 315 320  
 Phe Ser Leu Pro Ile Asp Ala Glu Val Met Ser Glu Ile Arg Leu Gln  
 325 330 335  
 Ala

<210> 87  
 <211> 1062  
 <212> DNA  
 <213> Unknown

<220>



<223> Obtained from an environmental sample

<400> 87

```

atggcggaat cgaagctgaa ggtcgccgca attcaagttg cgcccggtgtt catggatcgc      60
gatgccacga tcgcccgcgc ctgcgagcgg atcgccgaag ccgcccgcgc cggcgcgagg      120
ttggtggtct ttcccagggc attcgtgccg gggatatccc actggatctg ggtggcgcgg      180
ccaagccaac gcaaactgct caatgatctt tacgcgcacc tcgtctcgca gtcggtcgac      240
gtgccgtcgg cctccgtgga tcgtttgcgc gacgcggctc gcgacggcgg ggtcacggtg      300
gtgatcggcg tcaacgagcg caacaccgaa gcgagcggcg cgagcctcta caacaccgcg      360
ctcgtgatcg gtccactggg gcagctgata ggccgccacc gcaagcttgt gccgaccggg      420
ccggagcgca tgggtgtggg gcagggcgac ggcagcacgc tcgacgtcta cgacacaccc      480
gtcggcaagc ttctgacgtt gatctgctgg gagaactaca tgccgctcgc gcgctacgcc      540
atggcgggcg ggggcgcgcg catccacgtc gccggcacgt gggaccgcgg cgagccgtgg      600
atctcgacca tgcgtcatgt ggcgacggag ggccgcgtat tcgtgattag ctgttgcatt      660
gcgctgcgca aacgagacat tcccgcggag ctcgagttcg cgatgctcta tcccgcggg      720
cgcgaaatgga tcaacgccgg tgattcgctg gtcgtgaatc ccgctggcca gatcatcgct      780
gggccgttgc acgagcagga aggaatcctc tacgccgagc tcgagcgcaa tcagatgacc      840
ggtccgcgtt ggatgttcga cgccgccggc cattacgcgc gaccggacgt cttccaactc      900
acggtaaacc gtcccccgcg cccgatgctg cgggagggcg gggcaaagac gagtgaggca      960
aacacgagag atgccgtacc catggacagc acgccctcga gatcgcgggc ccgcgcgggt      1020
gcgcgaaagg ccgcacgcac cggtcgctcc aagcggcggt ga      1062

```

<210> 88

<211> 353

<212> PRT

<213> Unknown

<220>

<223> Obtained from an environmental sample

<400> 88

```

Met Ala Glu Ser Lys Leu Lys Val Ala Ala Ile Gln Val Ala Pro Val
 1          5          10          15
Phe Met Asp Arg Asp Ala Thr Ile Ala Arg Ala Cys Glu Arg Ile Ala
 20          25          30
Glu Ala Ala Arg Ala Gly Ala Glu Leu Val Val Phe Pro Glu Ala Phe
 35          40          45
Val Pro Gly Tyr Pro Asp Trp Ile Trp Val Ala Arg Pro Ser Gln Arg
 50          55          60
Lys Leu Leu Asn Asp Leu Tyr Ala His Leu Val Ser Gln Ser Val Asp
 65          70          75          80
Val Pro Ser Ala Ser Val Asp Arg Leu Arg Asp Ala Ala Arg Asp Gly
 85          90          95
Gly Val Thr Val Val Ile Gly Val Asn Glu Arg Asn Thr Glu Ala Ser
 100         105         110
Gly Ala Ser Leu Tyr Asn Thr Ala Leu Val Ile Gly Pro Leu Gly Gln
 115         120         125
Leu Ile Gly Arg His Arg Lys Leu Val Pro Thr Gly Pro Glu Arg Met
 130         135         140
Val Trp Ala Gln Gly Asp Gly Ser Thr Leu Asp Val Tyr Asp Thr Pro
 145         150         155         160
Val Gly Lys Leu Ser Thr Leu Ile Cys Trp Glu Asn Tyr Met Pro Leu
 165         170         175
Ala Arg Tyr Ala Met Ala Ala Trp Gly Ala Arg Ile His Val Ala Gly
 180         185         190
Thr Trp Asp Arg Gly Glu Pro Trp Ile Ser Thr Met Arg His Val Ala
 195         200         205
Thr Glu Gly Arg Val Phe Val Ile Ser Cys Cys Met Ala Leu Arg Lys
 210         215         220
Arg Asp Ile Pro Ala Glu Leu Glu Phe Ala Met Leu Tyr Pro Asp Gly
 225         230         235         240

```







|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Ile | Pro | Asp | Ser | Thr | Thr | Gln | Lys | Leu | Cys | Gln | Ala | Ala | Lys | Asp | Ala |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Gly | Val | Tyr | Val | Ala | Val | Gly | Ile | His | Glu | Arg | Asn | Ala | Glu | Ala | Ser |
|     |     |     | 100 |     |     |     | 105 |     |     |     |     |     | 110 |     |     |
| Gly | Phe | Thr | Leu | Phe | Asn | Thr | Leu | Leu | Tyr | Ile | Asn | Asp | Gln | Gly | Ser |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Ile | Ile | Gly | Lys | His | Arg | Lys | Leu | Ile | Pro | Thr | Gly | Gly | Glu | Arg | Leu |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Val | Trp | Gly | Gln | Gly | Asn | Gly | Asp | Thr | Leu | Ala | Ala | Phe | Asp | Thr | His |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Phe | Gly | Lys | Leu | Gly | Gly | Leu | Leu | Cys | Trp | Glu | Asn | Tyr | Met | Pro | Leu |
|     |     |     | 165 |     |     |     |     | 170 |     |     |     |     |     | 175 |     |
| Ala | Arg | Gln | Ala | Met | Tyr | Ala | Val | Gly | Thr | Glu | Val | Tyr | Val | Ala | Pro |
|     |     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |     |
| Thr | Trp | Asp | Ser | Ser | Glu | Asn | Trp | Leu | Leu | Ser | Met | Arg | His | Ile | Ala |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Arg | Glu | Gly | Gly | Met | Phe | Val | Ile | Asn | Val | Cys | Gln | Ala | Val | Arg | Lys |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Asp | Asp | Ile | Pro | Asp | Arg | Tyr | Ala | Phe | Lys | Gln | Leu | Tyr | Ser | Gly | Asn |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Ser | Glu | Trp | Ile | Asn | Ser | Gly | Asn | Ser | Cys | Ile | Ile | Asn | Pro | Arg | Gly |
|     |     |     | 245 |     |     |     |     | 250 |     |     |     |     |     | 255 |     |
| Glu | Ile | Ile | Ala | Gly | Pro | Ser | Ser | Asn | Arg | Gln | Glu | Ile | Leu | Tyr | Ala |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Asp | Leu | Asp | Leu | Ser | Leu | Ile | Thr | Lys | Ser | Lys | Arg | Met | Phe | Asp | Val |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Thr | Gly | His | Tyr | Ala | Arg | Pro | Asp | Val | Phe | Arg | Tyr | Glu | Ile | Lys | Lys |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Ser |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 305 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

<210> 91  
 <211> 939  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

|            |                                                             |
|------------|-------------------------------------------------------------|
| <400> 91   |                                                             |
| atgaccaaaa | tcgctgtcat tcaagaacct ccggtctatc tgaatctgag taaatcgatg 60   |
| gacagagcgg | tcgacttgat tgccaatgct gcaagcaagg ggtgtgagtt gattgtgttt 120  |
| cccgaagcct | ggcttgcagg ttaccccacc ttcgtctggc gtcttgcgcc gggcagcgga 180  |
| atgggaaaaa | ctgatgagct ttacgcgcgt ttgctcgcca actcggtcga ccgtagcaaa 240  |
| gaggggctta | gaccattgca ggaggccgca aaggagcatg gcgttgatcat tgtgctgggt 300 |
| tatcaagagg | tggatggcgc gggaagcagc agcacgatct tcaacagctg tgcgattatt 360  |
| gatgcggacg | ggcgactggc caacaatcat cgcaagtga tgcccaccaa tccggagagg 420   |
| atggtttggg | gttttggcga cggttcaggc ctgaacgtcg ttgacaccgc ggtgggcagg 480  |
| atcggcacgc | tgatttgctg ggaaaactac atgccgttag cgcgctacgc gctgtatgtc 540  |
| caaaacatcg | aaatctatgt tgccccgact tgggacagtg gtgccatgtg gcaggcgacc 600  |
| ctgcagcata | tcgcgcgcga aggtggctgc tgggtcatcg gatgtgcaac gtcgctggaa 660  |
| gcctctgaca | tcccggacga cgttcccat cgggatgagc tattcccga caaagacgaa 720    |
| tgggtaaacc | ctggcgatgc ggtggtttat aagccatttg gcggcattgt ggccggcccc 780  |
| atgcatcagg | aaaaggggct tctcatcgca gagttggacg tcgccgctgt tcagtcgtca 840  |
| cgtcggaagt | tcgatgagc cgggcactac gctcgccccg atgtcttcaa actgcatgtg 900   |
| aatcgacccg | cgatgcggcc agttgatttc acgaattag 939                         |

<210> 92  
 <211> 312  
 <212> PRT



WO 03/097810

PCT/US03/15712

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 92

```

Met Thr Lys Ile Ala Val Ile Gln Glu Pro Pro Val Tyr Leu Asn Leu
 1          5          10          15
Ser Lys Ser Met Asp Arg Ala Val Asp Leu Ile Ala Asn Ala Ala Ser
 20          25          30
Lys Gly Cys Glu Leu Ile Val Phe Pro Glu Ala Trp Leu Ala Gly Tyr
 35          40          45
Pro Thr Phe Val Trp Arg Leu Ala Pro Gly Ser Gly Met Gly Lys Thr
 50          55          60
Asp Glu Leu Tyr Ala Arg Leu Leu Ala Asn Ser Val Asp Arg Ser Lys
 65          70          75          80
Glu Gly Leu Arg Pro Leu Gln Glu Ala Ala Lys Glu His Gly Val Val
 85          90          95
Ile Val Leu Gly Tyr Gln Glu Val Asp Gly Ala Gly Ser Ser Ser Thr
100          105          110
Ile Phe Asn Ser Cys Ala Ile Ile Asp Ala Asp Gly Arg Leu Ala Asn
115          120          125
Asn His Arg Lys Leu Met Pro Thr Asn Pro Glu Arg Met Val Trp Gly
130          135          140
Phe Gly Asp Gly Ser Gly Leu Asn Val Val Asp Thr Ala Val Gly Arg
145          150          155          160
Ile Gly Thr Leu Ile Cys Trp Glu Asn Tyr Met Pro Leu Ala Arg Tyr
165          170          175
Ala Leu Tyr Val Gln Asn Ile Glu Ile Tyr Val Ala Pro Thr Trp Asp
180          185          190
Ser Gly Ala Met Trp Gln Ala Thr Leu Gln His Ile Ala Arg Glu Gly
195          200          205
Gly Cys Trp Val Ile Gly Cys Ala Thr Ser Leu Glu Ala Ser Asp Ile
210          215          220
Pro Asp Asp Val Pro His Arg Asp Glu Leu Phe Pro Asn Lys Asp Glu
225          230          235          240
Trp Val Asn Pro Gly Asp Ala Val Val Tyr Lys Pro Phe Gly Gly Ile
245          250          255
Val Ala Gly Pro Met His Gln Glu Lys Gly Leu Leu Ile Ala Glu Leu
260          265          270
Asp Val Ala Ala Val Gln Ser Ser Arg Arg Lys Phe Asp Ala Ser Gly
275          280          285
His Tyr Ala Arg Pro Asp Val Phe Lys Leu His Val Asn Arg Thr Ala
290          295          300
Met Arg Pro Val Asp Phe Thr Asn
305          310

```

&lt;210&gt; 93

&lt;211&gt; 978

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 93

```

atgcccacatca tcaaagccgc tgccgtgcaa atcagcccgg tgctttacag tcgcgaaggc 60
accgtggaca aggtctgtca acagatcatc gacctcggtc ggcaaggcgt gcagttcgcc 120
gtctttccgg aaacggtggt gccttactac ccgtactttt cgtttgtgca gccggccttt 180
gccatgggcg cacagcacct caagttgctg gatcaatcgg tgacagtgcc gtcggccgcc 240
accttgcca tcggtgaagc ttgcaagcaa gcagggatag tggtgtccat cggcgtcaac 300

```



```

gaacgcgatg gcggtacgat ctacaacgcg caattactct tcgatgccga cggcagcctg 360
attcagcatc gccgcaaaat caccocgacc tatcacgaac gcatgggtctg ggggcaaggc 420
gatggttccg gcctgcgcgc catcgacagt gcagtggggc gcattggctc cctggcctgt 480
tgaggagcatt acaacccgct ggctcgttat gccttgatgg ccgatggcga gcagatccac 540
gccgcgatgt ttcccggctc gctgggtgggc gacatttttg ccgagcagat cgaagtcacc 600
atccgccatc acgccttgga gtccggctgt ttctgtgtca acgccaccgc ctggctggac 660
gccgatcagc agggccaaat catgcaagac accggttgca gcctcggccc gatctcgggt 720
ggctgcttca ccgccatcgt ttcccctgaa ggcaagttgc tcggtgagcc gctgcgttcc 780
ggcgaagggg tggatgatcg cgatctcgat ctggcactga tcgataagcg taaacggatg 840
atggattcgg tcgggcatta cagtcgcccg gaactgctca gcctgttgat cgaccgcacg 900
cccacagcgc atgtgcatga acgcagcgcg cacctgggtg ctgtcgctac cgaggagttc 960
gatcatgcaa accaatga 978

```

&lt;210&gt; 94

&lt;211&gt; 325

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 94

```

Met Pro Ile Ile Lys Ala Ala Ala Val Gln Ile Ser Pro Val Leu Tyr
 1          5          10          15
Ser Arg Glu Gly Thr Val Asp Lys Val Cys Gln Gln Ile Ile Asp Leu
          20          25          30
Gly Arg Gln Gly Val Gln Phe Ala Val Phe Pro Glu Thr Val Val Pro
          35          40          45
Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Pro Ala Phe Ala Met Gly Ala
          50          55          60
Gln His Leu Lys Leu Leu Asp Gln Ser Val Thr Val Pro Ser Ala Ala
65          70          75          80
Thr Leu Ala Ile Gly Glu Ala Cys Lys Gln Ala Gly Ile Val Val Ser
          85          90          95
Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Ile Tyr Asn Ala Gln Leu
          100          105          110
Leu Phe Asp Ala Asp Gly Ser Leu Ile Gln His Arg Arg Lys Ile Thr
          115          120          125
Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp Gly Ser Gly
          130          135          140
Leu Arg Ala Ile Asp Ser Ala Val Gly Arg Ile Gly Ser Leu Ala Cys
145          150          155          160
Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met Ala Asp Gly
          165          170          175
Glu Gln Ile His Ala Ala Met Phe Pro Gly Ser Leu Val Gly Asp Ile
          180          185          190
Phe Ala Glu Gln Ile Glu Val Thr Ile Arg His His Ala Leu Glu Ser
          195          200          205
Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala Asp Gln Gln
          210          215          220
Gly Gln Ile Met Gln Asp Thr Gly Cys Ser Leu Gly Pro Ile Ser Gly
225          230          235          240
Gly Cys Phe Thr Ala Ile Val Ser Pro Glu Gly Lys Leu Leu Gly Glu
          245          250          255
Pro Leu Arg Ser Gly Glu Gly Val Val Ile Ala Asp Leu Asp Leu Ala
          260          265          270
Leu Ile Asp Lys Arg Lys Arg Met Met Asp Ser Val Gly His Tyr Ser
          275          280          285
Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Thr Pro Thr Ala His
          290          295          300
Val His Glu Arg Ser Ala His Leu Val Ala Val Ala Thr Glu Glu Phe

```



320

| <400> 96 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Met      | Ser | Asn | Glu | Asn | Asn | Ile | Ala | Thr | Phe | Lys | Val | Ala | Ala | Val | Gln |  |
| 1        |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |  |
| Ala      | Thr | Pro | Val | Phe | Leu | Asp | Arg | Glu | Ala | Thr | Ile | Asp | Lys | Ala | Cys |  |
|          |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |  |
| Ala      | Leu | Ile | Ala | Thr | Ala | Gly | Ser | Glu | Gly | Ala | Arg | Leu | Ile | Val | Phe |  |
|          |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |  |
| Pro      | Glu | Ala | Phe | Ile | Pro | Thr | Tyr | Pro | Glu | Trp | Val | Trp | Gly | Ile | Pro |  |
|          | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |  |
| Ser      | Gly | Glu | Gln | Gly | Leu | Leu | Asn | Glu | Leu | Tyr | Ala | Glu | Leu | Leu | Thr |  |
| 65       |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |  |
| Asn      | Ala | Val | Thr | Ile | Pro | Ser | Asp | Ala | Thr | Asp | Arg | Leu | Cys | Glu | Ala |  |
|          |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |  |
| Ala      | Gln | Leu | Ala | Asn | Ala | Tyr | Val | Val | Met | Gly | Met | Ser | Glu | Arg | Asn |  |
|          |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |
| Val      | Glu | Ala | Ser | Gly | Ala | Ser | Leu | Tyr | Asn | Thr | Leu | Leu | Tyr | Ile | Asn |  |
|          |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |
| Ala      | Gln | Gly | Glu | Ile | Leu | Gly | Lys | His | Arg | Lys | Leu | Val | Pro | Thr | Gly |  |
|          | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |
| Gly      | Glu | Arg | Leu | Val | Trp | Ala | Gln | Gly | Asp | Gly | Ser | Thr | Leu | Gln | Val |  |
| 145      |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |  |
| Tyr      | Asp | Thr | Pro | Leu | Gly | Lys | Leu | Gly | Gly | Leu | Ile | Cys | Trp | Glu | Asn |  |



```

                165                170                175
Tyr Met Pro Leu Ala Arg Tyr Ala Met Tyr Ala Trp Gly Thr Gln Ile
                180                185                190
Tyr Val Ala Ala Thr Trp Asp Arg Gly Gln Pro Trp Leu Ser Thr Leu
                195                200                205
Arg His Ile Ala Lys Glu Gly Arg Val Tyr Val Ile Gly Cys Cys Ile
                210                215                220
Ala Met Arg Lys Asp Asp Ile Pro Asp Arg Tyr Thr Met Lys Gln Lys
225                230                235                240
Tyr Tyr Ala Glu Met Asp Glu Trp Met Asn Val Gly Asp Ser Val Ile
                245                250                255
Val Asn Pro Glu Gly His Phe Ile Ala Gly Pro Val Arg Lys Gln Glu
                260                265                270
Glu Ile Leu Tyr Ala Glu Ile Asp Pro Arg Met Val Gln Gly Pro Lys
                275                280                285
Trp Met Leu Asp Val Ala Gly His Tyr Ala Arg Pro Asp Val Phe Gln
                290                295                300
Leu Thr Val His Thr Asp Val Arg Arg Met Met Arg Val Glu Asp Asp
305                310                315                320
Ser

```

<210> 97  
 <211> 1017  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 97
atgggcatcg aacatccgaa atacaagggtc gccgtggtgc aagctgcgcc cgcctggctc      60
gacctcgacg cgtcgatcga caagacgatc gggctgatcg aggaggcggc gaagaaaggc      120
gccaaagtga tcgctttccc cgaagccttc attcccggct acccttggca catctggctc      180
gactcaccgc cctgggcgat cggccgcggt ttcgtgcagc gctatttcga caattcgctc      240
gcctacgaca gcccacaggc ggaaaggctg cgacaggccg tgcggaaggc caagctcacc      300
gccgtgatcg gcctgtccga gcgcgacggc ggcagcctct atctcgcgca gtggctgatc      360
gggcccgcag gtgagaccat cgcaaagcgc cgcaagctgc ggccgacca tgccgagcgc      420
accgtctatg gcgaaggcga cggcagcgat ctgcgcgtcc atgagcgggc cgacatcggc      480
cggctcggcg cgctgtgctg ctgggagcat ctgcagccgc tgcgaaatt cgccatgtac      540
gcccagaacg agcaggtaca tgtcgcggcc tggccgagct tctcgctcta cgatcccttc      600
gcgcctgcgc tgggcgcgga ggtgaacaac gccgcctccc gcatctatgc ggtggaaggc      660
tcctgcttcg tgctcgcacc gtgcgcgacg gtctcgcagg ccatgatcga cgagctctgc      720
gatcggccgg acaagcacgc gctgctgcat gccggcggcg gcttcgccgc gatctacggg      780
cccgacggca gccagatcgg cgacaagctg ccgcccgagc aggagggcct gctgatcgcc      840
gagatcgatc tgggcgcgat cggcgtcgcc aagaacgcgg ccgatcccgc cgggcattat      900
tcgcggcccg acgtcacgcg gctcctgctc aacaggaagc cgaacaagcg cgtggagcag      960
ttcgcgctgc ccgtcgcacac ggtcgagccc gtcgacgtcg cggcggcagc aagctga      1017

```

<210> 98  
 <211> 338  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 98
Met Gly Ile Glu His Pro Lys Tyr Lys Val Ala Val Val Gln Ala Ala
  1                5                10                15
Pro Ala Trp Leu Asp Leu Asp Ala Ser Ile Asp Lys Thr Ile Gly Leu

```







```

gaagtgacca tccggcatca cgcactggaa tccggctgct tcgtgatcaa ttccaccggc 660
tggtgaccg agccgcagat cgagtcgatc acgaaagatc cgggcctgca gaaggcgctt 720
cgcggcggct gcaacacggc gatcatctcg cccgaaggcc agcatctcgc cccgccgctg 780
cgtgagggcg agggcatggt catcgctgac ctggacatgt cgctgatcac caaacgcaaa 840
cgcatgatgg attctgtcgg ccactacgcg cggcccgaac tgctgagcct cgccatcaac 900
gaccggcccg cggtcacgtc ggcacccatg aacagcttct catcttcaac cgggggattg 960
caccttgaac gcgaacgaga ccttgtcggc cgtgagccgg caattgatga ctga 1014

```

&lt;210&gt; 100

&lt;211&gt; 337

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 100

```

Met Pro Asp Lys Arg Ile Val Arg Ala Ala Ala Val Gln Ile Ala Pro
 1          5          10          15
Asp Leu Glu Arg Pro Gly Gly Thr Leu Glu Lys Val Leu Glu Thr Ile
          20          25          30
Asp Asp Ala Arg Gln Gly Val Gln Leu Ile Val Phe Pro Glu Thr
          35          40          45
Phe Leu Pro Tyr Tyr Pro Tyr Phe Ser Phe Val Arg Ala Pro Val Ala
          50          55          60
Ser Gly Ala Glu His Met Arg Leu Tyr Asp Glu Ala Val Val Val Pro
65          70          75          80
Gly Pro Val Thr His Ala Val Ala Glu Arg Ala Arg Arg His Gly Met
          85          90          95
Val Val Val Leu Gly Val Asn Glu Arg Asp His Gly Ser Leu Tyr Asn
          100          105          110
Ala Gln Leu Ile Phe Asp Thr Asp Gly Glu Leu Leu Leu Lys Arg Arg
          115          120          125
Lys Ile Thr Pro Thr Phe His Glu Arg Met Ile Trp Gly Met Gly Asp
          130          135          140
Ala Ala Gly Leu Lys Val Ala Glu Thr Arg Ile Gly Arg Val Gly Ala
145          150          155          160
Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met
          165          170          175
Thr Gln His Glu Glu Ile His Cys Ser Gln Phe Pro Gly Ser Leu Val
          180          185          190
Gly Pro Ile Phe Gly Glu Gln Ile Glu Val Thr Ile Arg His His Ala
          195          200          205
Leu Glu Ser Gly Cys Phe Val Ile Asn Ser Thr Gly Trp Leu Thr Glu
          210          215          220
Pro Gln Ile Glu Ser Ile Thr Lys Asp Pro Gly Leu Gln Lys Ala Leu
225          230          235          240
Arg Gly Gly Cys Asn Thr Ala Ile Ile Ser Pro Glu Gly Gln His Leu
          245          250          255
Ala Pro Pro Leu Arg Glu Gly Glu Gly Met Val Ile Ala Asp Leu Asp
          260          265          270
Met Ser Leu Ile Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly His
          275          280          285
Tyr Ala Arg Pro Glu Leu Leu Ser Leu Ala Ile Asn Asp Arg Pro Ala
          290          295          300
Val Thr Ser Ala Pro Met Asn Ser Phe Ser Ser Ser Thr Gly Gly Leu
305          310          315          320
His Leu Glu Arg Glu Arg Asp Leu Val Gly Arg Glu Pro Ala Ile Asp
          325          330          335
Asp

```



WO 03/097810

PCT/US03/15712

<210> 101  
 <211> 1065  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 101  
 atggcgacag tccatccgaa atttaaagta gccgcccgtcc aggcggccccc ggcctttctc 60  
 gacctcgacg cgtcgggtgga aaaagcgggtg cgcctgattg atgaagccgg cgccgctggt 120  
 gcccggtctca tcgcgtttcc agagactttt atccccgggt atccgtggtg gatctggctc 180  
 ggtgctccgg cctgggagat catgcgcggc ttcgtctccc gctatttcga caactcgctg 240  
 cagtacggca ccccggaagc cgaccggctg cgggcagccg ccaaacgcaa caaatgttc 300  
 gtcgcgctcg gactgtcaga gcgcgacggc ggcagtctct acatcgccca atggattatc 360  
 ggacccgacg gcgagacggg cgcaacgcgc cgcaagctca agcctactca cgccgagcgg 420  
 acgggtgttcg gcgaaggcga tggctcgcac cttgcgggtc acgaacttga tatcgggcgg 480  
 gtcgggtgcgc tgtgctgttg ggagcacctg cagccactgt cgaagtacgc gatgtatgcg 540  
 cagaacgagc aagttcatat cgcggcgttg ccgagctttt cgctttacga tccgttcgcg 600  
 catgcgcttg gcgccgaggt caacaacgcg gcgagcaaga tctacgcggg cgaaggctca 660  
 tgctttgtga ttgcgccatg cgcgaccgtt tcccaggcga tgatcgacga attgtgtgac 720  
 tcgcccgaga agcatcagtt cctgcacgtc ggcggcgggt tcgccgtgat ctatgggtccc 780  
 gacggcgcgc cactcgccaa gccactggcg cccgatcagg aggggtctct ttacgcggat 840  
 atcgacctcg gcatgatttc ggtcgcgaaa gcggcggccg atccggctgg acattacgcg 900  
 cgcccggacg tgaccctctt gttgttcaac aatcgtcctg ggaaccgggt ggagacactc 960  
 gcgctgccgg tcgaccagga ggcagaggcg ggagcaggcg gcaaacctgc gcccaagtca 1020  
 ccgagtgtcg ctgcgttcac actgacgcag gcggcagccg agtag 1065

<210> 102  
 <211> 354  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 102  
 Met Ala Thr Val His Pro Lys Phe Lys Val Ala Ala Val Gln Ala Ala  
 1 5 10 15  
 Pro Ala Phe Leu Asp Leu Asp Ala Ser Val Glu Lys Ala Val Arg Leu  
 20 25 30  
 Ile Asp Glu Ala Gly Ala Ala Gly Ala Arg Leu Ile Ala Phe Pro Glu  
 35 40 45  
 Thr Phe Ile Pro Gly Tyr Pro Trp Trp Ile Trp Leu Gly Ala Pro Ala  
 50 55 60  
 Trp Ala Ile Met Arg Gly Phe Val Ser Arg Tyr Phe Asp Asn Ser Leu  
 65 70 75 80  
 Gln Tyr Gly Thr Pro Glu Ala Asp Arg Leu Arg Ala Ala Ala Lys Arg  
 85 90 95  
 Asn Lys Met Phe Val Ala Leu Gly Leu Ser Glu Arg Asp Gly Gly Ser  
 100 105 110  
 Leu Tyr Ile Ala Gln Trp Ile Ile Gly Pro Asp Gly Glu Thr Val Ala  
 115 120 125  
 Thr Arg Arg Lys Leu Lys Pro Thr His Ala Glu Arg Thr Val Phe Gly  
 130 135 140  
 Glu Gly Asp Gly Ser His Leu Ala Val His Glu Leu Asp Ile Gly Arg  
 145 150 155 160  
 Val Gly Ala Leu Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys Tyr  
 165 170 175  
 Ala Met Tyr Ala Gln Asn Glu Gln Val His Ile Ala Ala Trp Pro Ser







Val Tyr Phe Asp Arg Glu Ala Ser Thr Glu Lys Ala Arg Gly Leu Ile  
 20 25 30  
 Arg Glu Ala Gly Glu Lys Gly Val Asp Leu Leu Ala Phe Gly Glu Thr  
 35 40 45  
 Trp Leu Thr Gly Tyr Pro Tyr Trp Lys Asp Ala Pro Trp Ser Arg Glu  
 50 55 60  
 Tyr Asn Asp Leu Arg Ala Arg Tyr Val Ala Asn Gly Val Met Ile Pro  
 65 70 75 80  
 Gly Pro Glu Thr Asp Ala Leu Cys Gln Ala Ala Ala Glu Ala Gly Val  
 85 90 95  
 Asp Val Ala Ile Gly Val Val Glu Leu Glu Pro Gly Ser Leu Ser Ser  
 100 105 110  
 Val Tyr Cys Thr Leu Leu Phe Ile Ser Arg Glu Gly Glu Ile Leu Gly  
 115 120 125  
 Arg His Arg Lys Leu Lys Pro Thr Asp Ser Glu Arg Arg Tyr Trp Ser  
 130 135 140  
 Glu Gly Asp Ala Thr Gly Leu Arg Val Tyr Glu Arg Pro Tyr Gly Arg  
 145 150 155 160  
 Leu Ser Gly Leu Asn Cys Trp Glu His Leu Met Met Leu Pro Gly Tyr  
 165 170 175  
 Ala Leu Ala Ala Gln Gly Thr Gln Phe His Val Ala Ala Trp Pro Asn  
 180 185 190  
 Met Ala Ser Ser Ala Ser Glu Leu Leu Ser Arg Ala Tyr Ala Tyr Gln  
 195 200 205  
 Ala Gly Cys Tyr Val Leu Cys Ala Gly Gly Leu Gly Pro Ala Pro Gly  
 210 215 220  
 Glu Leu Pro Asp Gly Ile Ala Ala Glu Ser Leu Asp His Leu Thr Gly  
 225 230 235 240  
 Glu Ser Cys Ile Ile Asp Pro Trp Gly Lys Val Ile Ala Gly Pro Val  
 245 250 255  
 Ser Cys Glu Glu Thr Leu Ile Thr Ala Arg Val Ser Thr Ala Ser Ile  
 260 265 270  
 Tyr Arg Arg Lys Ser Leu Thr Asp Val Gly Gly His Tyr Ser Arg Pro  
 275 280 285  
 Asp Val Phe Arg Phe Glu Val Asp Arg Ser Glu Arg Pro Arg Val Val  
 290 295 300  
 Phe Arg Asp Gly Asp Val Asp Asp Arg Gly  
 305 310

&lt;210&gt; 105

&lt;211&gt; 975

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 105

|            |            |            |             |            |            |     |
|------------|------------|------------|-------------|------------|------------|-----|
| atgaccattg | tcaaagccgc | tgccgtccag | attgcccccg  | ttctctacag | ccgtgaaggc | 60  |
| actgtagaaa | aggtcgtaa  | caagattcgc | gaactcggcg  | agaagggcgt | gcagttcgcc | 120 |
| gttttccctg | aaaccgtcgt | accgtactac | ccgtactttt  | cctttgtgca | gagcccttc  | 180 |
| aaaatgggtt | ccgagcacta | caaattgctc | gaccaggccg  | ttgtcgtgcc | gtcggcgacc | 240 |
| accgatgcca | tcggcaaagc | ggccaaggaa | gccaacatgg  | tggtgtccat | cggcgtcaac | 300 |
| gaacgcgatg | gcagcaccct | ctacaacacg | cagttgctgt  | ttgatgccga | cggcactttg | 360 |
| attcaggccc | gtcgcaagat | ttcaccgacc | taccacgaac  | gcatgatctg | gggcatgggc | 420 |
| gacggttccg | gcctgcgcgc | caccgacagc | gcggtcgggc  | gcatcgga   | attggcctgc | 480 |
| tggaacatt  | acaatccgct | ggcgcgttac | gccttgatcg  | aagacggcga | acagatccac | 540 |
| gcctcgatgt | acccgggctc | gttcgcaggt | cctttattca  | ctcgccagat | ggaagtcagc | 600 |
| atccgcatgc | atgccctgga | atcggcgtgc | ttcgtgggtca | actcgaccgc | gtggttgtag | 660 |
| ccggaacagc | aagcccagat | catggccgac | accggttgcg  | agatcggggc | gatctccggc | 720 |
| ggctgctaca | ccgcgatcat | cgaccacag  | ggtgaagtcg  | tcggcgcact | gaccgaaggc | 780 |



|       |         |
|-------|---------|
| <210> | 107     |
| <211> | 981     |
| <212> | DNA     |
| <213> | Unknown |



&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 107

```

atggccatca ttgcgcgagc agccgtacag atcagcccgg ttctttacag ccgcgaaggc      60
accgtggaca aggtctgcca gcagatcatc acccttggca aacaggggtgt gcagttcgcc      120
gtgttcccgg aaacgggtgt gccgtactac ccctatTTTT cctttgtgca gccggcggtc      180
gccatgggtg cgcaacacct caaattgcta gatcaatctg taaccgtgcc atcggccgcc      240
accctggcga ttggcgaagc gtgcaagcaa gcaggaatgg tcgtttccat cggagtcaat      300
gaacgcgatg gcggtacgat ttacaacgcg caattactct tcgatgctga cggcacgctg      360
attcagcatc ggcgcaaaat caccocgacc taccacgagc gcatgggtctg ggggcagggc      420
gatggttccg gtctgcgcgc catcgacagc gcggtcgggc gcatcggtct cctggcatgc      480
tggaacatt acaacccgct ggcccgttac gccttgatgg cagacggcga acagatccac      540
gccgcgatgt ttcccgggtc cctgggtgggt gacatcttcg ccgagcagat cgaggtcacc      600
atccgccatc acgcattgga gtcaggatgc ttcgtgggtca atgcaacagc ctggctggat      660
gcggatcagc agggccaaat aatgcaggac acaggttgcg gccttggtcc catctcgggc      720
ggctgcttca ccgcgatcgt atcgccggaa gggaagctac ttggagagcc gcttcgctcc      780
ggggaaggcg tagtgattgc cgacctcgat acggccttga tcgacaagcg caaacggatg      840
atggattcag taggtcatta cagtcgtccc gagctgctca gcctattgat cgatcgatcg      900
ccgactgcgc atgttcatga acgcgccggc tttgtttcga gcaacgccgg tttgcaggag      960
gtcgcccatg cagaccaatg a                                     981

```

&lt;210&gt; 108

&lt;211&gt; 326

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 108

```

Met Ala Ile Ile Arg Ala Ala Ala Val Gln Ile Ser Pro Val Leu Tyr
 1          5          10          15
Ser Arg Glu Gly Thr Val Asp Lys Val Cys Gln Gln Ile Ile Thr Leu
          20          25          30
Gly Lys Gln Gly Val Gln Phe Ala Val Phe Pro Glu Thr Val Val Pro
          35          40          45
Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Pro Ala Phe Ala Met Gly Ala
          50          55          60
Gln His Leu Lys Leu Leu Asp Gln Ser Val Thr Val Pro Ser Ala Ala
65          70          75          80
Thr Leu Ala Ile Gly Glu Ala Cys Lys Gln Ala Gly Met Val Val Ser
          85          90          95
Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Ile Tyr Asn Ala Gln Leu
          100          105          110
Leu Phe Asp Ala Asp Gly Thr Leu Ile Gln His Arg Arg Lys Ile Thr
          115          120          125
Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp Gly Ser Gly
          130          135          140
Leu Arg Ala Ile Asp Ser Ala Val Gly Arg Ile Gly Ser Leu Ala Cys
145          150          155          160
Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met Ala Asp Gly
          165          170          175
Glu Gln Ile His Ala Ala Met Phe Pro Gly Ser Leu Val Gly Asp Ile
          180          185          190
Phe Ala Glu Gln Ile Glu Val Thr Ile Arg His His Ala Leu Glu Ser
          195          200          205
Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala Asp Gln Gln
210          215          220
Gly Gln Ile Met Gln Asp Thr Gly Cys Gly Leu Gly Pro Ile Ser Gly

```



```

225          230          235          240
Gly Cys Phe Thr Ala Ile Val Ser Pro Glu Gly Lys Leu Leu Gly Glu
          245          250          255
Pro Leu Arg Ser Gly Glu Gly Val Val Ile Ala Asp Leu Asp Thr Ala
          260          265          270
Leu Ile Asp Lys Arg Lys Arg Met Met Asp Ser Val Gly His Tyr Ser
          275          280          285
Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Ser Pro Thr Ala His
          290          295          300
Val His Glu Arg Ala Gly Phe Val Ser Ser Asn Ala Gly Leu Gln Glu
305          310          315          320
Val Ala His Ala Asp Gln
          325

```

<210> 109  
 <211> 1092  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 109
atggccatca ttgcgcgagc agccgtacag atcagcccgg ttctttacag ccgcgaaggc      60
accgtggaca gggctctgcca gcagatcatc acccttgcca aacaagggtgt gcagttcgcc      120
gtgttcccgg aaacgggtggt gccgtactac ccctatTTTT cctttgtgca gccggcattt      180
gcgatgggtg cacaacacct caaattgctc gatcaatctg taaccgtgcc atcggccgcc      240
accctggcga ttggcgaagc gtgcaagcaa gcaggaatgg tcgtttccat cggcgtcaat      300
gaacgcgatg gcggtacgat ttacaacgcg caattactct tcgatgctga cggcactctg      360
attcagcatc ggcgcaaaat caccgacacc taccacgagc gcatgggtctg ggggcagggc      420
gatggttccg gtctgcgcgc catcgacagc gcggtcgggc gcatcggtc cctggcatgc      480
tgggaacatt acaacccgct ggcccgttac gccttgatgg cagacggcga acagatccac      540
gccgcgatgt ttcccggttc cctggtgggt gacatcttcg ccgagcagat cgaggtcacc      600
atccgccatc acgcattgga atcaggatgc ttctgtgtca atgcaacagc ttggctggat      660
gcggatcagc agggccaaat aatgcaggac acaggttgcg gccttggtcc catctcgggc      720
ggctgcttca ccgcgatcgt atcgccggaa gggaaagctac ttggagagcc gcttcgctca      780
ggggaaggcg tagtgattgc cgacctgat atggccttga tcgacaagcg caaacggatg      840
atggattcag taggtcatta cagtcgtccc gagctgctca gcctattgat cgatcgatcg      900
ccgactgcgc attttcatga acgcgccggg ctttgttccg agcgacgccg gtttgcagga      960
ggtcgcgcgt gcagaccaat gaattgctcg ctgacctgca aatccaaggc ctgcgttggc     1020
cggccgcgca aatggcttgt cgcgccaagg cggcgccggg ccttcagacc acaaggcgct     1080
gagcctaggt aa                                           1092

```

<210> 110  
 <211> 363  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

```

<400> 110
Met Ala Ile Ile Arg Ala Ala Ala Val Gln Ile Ser Pro Val Leu Tyr
  1          5          10          15
Ser Arg Glu Gly Thr Val Asp Arg Val Cys Gln Gln Ile Ile Thr Leu
          20          25          30
Gly Lys Gln Gly Val Gln Phe Ala Val Phe Pro Glu Thr Val Val Pro
          35          40          45
Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Pro Ala Phe Ala Met Gly Ala
          50          55          60
Gln His Leu Lys Leu Leu Asp Gln Ser Val Thr Val Pro Ser Ala Ala

```



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Thr | Leu | Ala | Ile | Gly | Glu | Ala | Cys | Lys | Gln | Ala | Gly | Met | Val | Val | Ser |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Ile | Gly | Val | Asn | Glu | Arg | Asp | Gly | Gly | Thr | Ile | Tyr | Asn | Ala | Gln | Leu |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Leu | Phe | Asp | Ala | Asp | Gly | Thr | Leu | Ile | Gln | His | Arg | Arg | Lys | Ile | Thr |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Pro | Thr | Tyr | His | Glu | Arg | Met | Val | Trp | Gly | Gln | Gly | Asp | Gly | Ser | Gly |
|     |     | 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Leu | Arg | Ala | Ile | Asp | Ser | Ala | Val | Gly | Arg | Ile | Gly | Ser | Leu | Ala | Cys |
| 145 |     |     |     |     | 150 |     |     |     | 155 |     |     |     |     |     | 160 |
| Trp | Glu | His | Tyr | Asn | Pro | Leu | Ala | Arg | Tyr | Ala | Leu | Met | Ala | Asp | Gly |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Glu | Gln | Ile | His | Ala | Ala | Met | Phe | Pro | Gly | Ser | Leu | Val | Gly | Asp | Ile |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Phe | Ala | Glu | Gln | Ile | Glu | Val | Thr | Ile | Arg | His | His | Ala | Leu | Glu | Ser |
|     |     | 195 |     |     |     |     | 200 |     |     |     | 205 |     |     |     |     |
| Gly | Cys | Phe | Val | Val | Asn | Ala | Thr | Ala | Trp | Leu | Asp | Ala | Asp | Gln | Gln |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Gly | Gln | Ile | Met | Gln | Asp | Thr | Gly | Cys | Gly | Leu | Gly | Pro | Ile | Ser | Gly |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Gly | Cys | Phe | Thr | Ala | Ile | Val | Ser | Pro | Glu | Gly | Lys | Leu | Leu | Gly | Glu |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Pro | Leu | Arg | Ser | Gly | Glu | Gly | Val | Val | Ile | Ala | Asp | Leu | Asp | Met | Ala |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Leu | Ile | Asp | Lys | Arg | Lys | Arg | Met | Met | Asp | Ser | Val | Gly | His | Tyr | Ser |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Arg | Pro | Glu | Leu | Leu | Ser | Leu | Leu | Ile | Asp | Arg | Ser | Pro | Thr | Ala | His |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Phe | His | Glu | Arg | Ala | Gly | Leu | Cys | Ser | Glu | Arg | Arg | Arg | Phe | Ala | Gly |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gly | Arg | Ala | Cys | Arg | Pro | Met | Asn | Cys | Ser | Leu | Thr | Cys | Lys | Ser | Lys |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Ala | Cys | Val | Gly | Arg | Pro | Arg | Lys | Trp | Leu | Val | Ala | Pro | Arg | Arg | Arg |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Arg | Ser | Phe | Arg | Pro | Gln | Gly | Ala | Glu | Pro | Arg |     |     |     |     |     |
|     |     | 355 |     |     |     |     | 360 |     |     |     |     |     |     |     |     |

<210> 111  
 <211> 990  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 111

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| atgcccaaaa | cagtacgtgc | cgcagcagtc | cagatcgcg  | ccgacctgac | gtcacgcgcc | 60  |
| ggcaccgtcg | agcgggtcct | caatgcaatc | gccgaagctg | ctgacaaagg | cgccgagctg | 120 |
| atcgtatttc | ccgagacctt | cgtgccctgg | tatccctatt | tcagtttcgt | tctgccacct | 180 |
| gtccagcaag | gccctgagca | tcttcgtctt | tatgaggaag | cagtcacggg | accatcagca | 240 |
| gaaacacggg | ccgtcgcgga | cgccgcgcgc | aaacgcaatg | cggttatcgt | ccttggcgct | 300 |
| aatgagcgcg | accacggctc | gctctataac | actcagctga | tcttcgacgc | ggatggcagc | 360 |
| ctgaaactca | agcgtcgcaa | gatcacgccg | acctatcacg | aacggatgat | ctggggccaa | 420 |
| ggcgatggcg | ccggcctgaa | ggttgtcgac | actgccgtcg | gtcgcgtggg | tgccctggca | 480 |
| tgctgggagc | attacaatcc | tctggcccgc | tatactttga | tggcccagca | tgaggaaatt | 540 |
| cacgcctctc | atttcccggg | ctcactgggc | ggcccgatat | tcggcgagca | aatcgaagtc | 600 |
| accatgcgcc | accacgcggt | ggaatcgggc | tgtttcgtgg | tcaatgccac | cggctggctg | 660 |
| agcgaggagc | agatcgcatc | tattcatccg | gaccccgctt | tgcaaaaggg | cctgcgcgat | 720 |
| ggctgcatga | cctgcatcat | cacgccggaa | ggacgccatg | tcgtaccgcc | gctgacctcg | 780 |
| ggcgaaggca | tcctgatcgg | cgatctggac | atgcggctca | ttaccaagcg | caagcggatg | 840 |



WO 03/097810

PCT/US03/15712

```

atggattcgg tcggacacta tgctcggcct gaactgctgc accttgtcca tgacacgacg      900
cccgacgcgc cacgcgagca ggtcggcctt tcaggcgatt ttcccgatgc ggagcaagac      960
aagctatttg aggaggttca taatgcgtga      990

```

<210> 112  
 <211> 329  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 112  
 Met Pro Lys Thr Val Arg Ala Ala Ala Val Gln Ile Ala Pro Asp Leu  
 1 5 10 15  
 Thr Ser Arg Ala Gly Thr Val Glu Arg Val Leu Asn Ala Ile Ala Glu  
 20 25 30  
 Ala Ala Asp Lys Gly Ala Glu Leu Ile Val Phe Pro Glu Thr Phe Val  
 35 40 45  
 Pro Trp Tyr Pro Tyr Phe Ser Phe Val Leu Pro Pro Val Gln Gln Gly  
 50 55 60  
 Pro Glu His Leu Arg Leu Tyr Glu Glu Ala Val Thr Val Pro Ser Ala  
 65 70 75 80  
 Glu Thr Arg Ala Val Ala Asp Ala Ala Arg Lys Arg Asn Ala Val Ile  
 85 90 95  
 Val Leu Gly Val Asn Glu Arg Asp His Gly Ser Leu Tyr Asn Thr Gln  
 100 105 110  
 Leu Ile Phe Asp Ala Asp Gly Ser Leu Lys Leu Lys Arg Arg Lys Ile  
 115 120 125  
 Thr Pro Thr Tyr His Glu Arg Met Ile Trp Gly Gln Gly Asp Gly Ala  
 130 135 140  
 Gly Leu Lys Val Val Asp Thr Ala Val Gly Arg Val Gly Ala Leu Ala  
 145 150 155 160  
 Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Thr Leu Met Ala Gln  
 165 170 175  
 His Glu Glu Ile His Ala Ser His Phe Pro Gly Ser Leu Val Gly Pro  
 180 185 190  
 Ile Phe Gly Glu Gln Ile Glu Val Thr Met Arg His His Ala Leu Glu  
 195 200 205  
 Ser Gly Cys Phe Val Val Asn Ala Thr Gly Trp Leu Ser Glu Glu Gln  
 210 215 220  
 Ile Ala Ser Ile His Pro Asp Pro Ala Leu Gln Lys Gly Leu Arg Asp  
 225 230 235 240  
 Gly Cys Met Thr Cys Ile Ile Thr Pro Glu Gly Arg His Val Val Pro  
 245 250 255  
 Pro Leu Thr Ser Gly Glu Gly Ile Leu Ile Gly Asp Leu Asp Met Arg  
 260 265 270  
 Leu Ile Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly His Tyr Ala  
 275 280 285  
 Arg Pro Glu Leu Leu His Leu Val His Asp Thr Thr Pro Ala Arg Ala  
 290 295 300  
 Arg Glu Gln Val Gly Leu Ser Gly Asp Phe Pro Asp Ala Glu Gln Asp  
 305 310 315 320  
 Lys Leu Phe Glu Glu Val His Asn Ala  
 325

<210> 113  
 <211> 993  
 <212> DNA  
 <213> Unknown



&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 113

```

atgacgaagg aacgcgcgcg gcgcagcctg cggcagctg ccatacagct tgaagccgaa      60
gtcggcgaca tcgccgcca tctcgcacgc atcgaggcga tggtcgagga ggctgcgggc      120
aagggcgccg aactgatcgc cattccggag ttctgcacct cccgcacatgcc cttcgatgca      180
cgcggtgcacg acgccgtgct gccgcgggac aacttcgtgg tcgatgcctt tcgccgcacg      240
gcagcgacgc acaactgccg gctcggcggc tccatgctca ttgccgacgg tggcgagatc      300
tacaaccgct accacttcgt cgaacccgac ggcagcgtgc atctgcacga caaggatctg      360
ccgacgatgt gggagaacgc cttctacacc ggcggtccg acgacggcgt cttcgacacc      420
ggcatcgggc gcgtcggcgc cgcggtgtgc tgggaactgg tacgcaccgg caccgtgcga      480
cgcatgctcg gtcgcgtcga cgtcgccatg accggcacgc attggtggac gatgcgcac      540
aactggggca gcgccgtcgc gcgcacgctg gccgcgatga cgcagtacaa ccgctacatg      600
tccgagaatg caccaccga attcgccgcg cgctgggtg tgccggtgct gcaggcctcg      660
cactgcggaa gcttcgcac cggtttcttg ctgctgccag gcagcgggcg tgcactgcc      720
tatgacaccg agtacgtcgg cgccacacag atcgtcgatg ccgatggcca catcctcgcc      780
caccgtcgca cgcaggaagg ccccggtgtc gtcgtcgccg acatcacgct cggtgccgcg      840
acgcccgagc tgccactgga agaccgcttc tggattcccg agctgccgct cttcctcaag      900
gcctactggc accaccagaa cctgtgcggc aagtcctact accgtcgcgt cggccgcgat      960
gccggcctgg cggcggcgga gcgttcggca tga      993

```

&lt;210&gt; 114

&lt;211&gt; 330

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 114

```

Met Thr Lys Glu Arg Ala Ala Arg Ser Leu Arg Ala Ala Ala Ile Gln
 1          5          10          15
Leu Glu Ala Glu Val Gly Asp Ile Ala Ala Asn Leu Ala Arg Ile Glu
 20          25          30
Ala Met Val Glu Glu Ala Ala Gly Lys Gly Ala Glu Leu Ile Ala Ile
 35          40          45
Pro Glu Phe Cys Thr Ser Arg Met Pro Phe Asp Ala Arg Val His Asp
 50          55          60
Ala Val Leu Pro Pro Asp Asn Phe Val Val Asp Ala Phe Arg Arg Met
 65          70          75          80
Ala Ala Thr His Asn Cys Arg Leu Gly Gly Ser Met Leu Ile Ala Asp
 85          90          95
Gly Gly Glu Ile Tyr Asn Arg Tyr His Phe Val Glu Pro Asp Gly Ser
100          105          110
Val His Leu His Asp Lys Asp Leu Pro Thr Met Trp Glu Asn Ala Phe
115          120          125
Tyr Thr Gly Gly Ser Asp Asp Gly Val Phe Asp Thr Gly Ile Gly Gly
130          135          140
Val Gly Ala Ala Val Cys Trp Glu Leu Val Arg Thr Gly Thr Val Arg
145          150          155          160
Arg Met Leu Gly Arg Val Asp Val Ala Met Thr Gly Thr His Trp Trp
165          170          175
Thr Met Pro His Asn Trp Gly Ser Ala Val Ala Arg Thr Leu Ala Ala
180          185          190
Met Thr Gln Tyr Asn Arg Tyr Met Ser Glu Asn Ala Pro Thr Glu Phe
195          200          205
Ala Arg Arg Leu Gly Val Pro Val Leu Gln Ala Ser His Cys Gly Ser
210          215          220
Phe Arg Thr Gly Phe Leu Leu Leu Pro Gly Ser Gly Arg Ala Leu Pro
225          230          235          240

```



Tyr Asp Thr Glu Tyr Val Gly Ala Thr Gln Ile Val Asp Ala Asp Gly  
 245 250 255  
 His Ile Leu Ala His Arg Arg Thr Gln Glu Gly Pro Gly Val Val Val  
 260 265 270  
 Ala Asp Ile Thr Leu Gly Ala Arg Thr Pro Glu Leu Pro Leu Glu Asp  
 275 280 285  
 Arg Phe Trp Ile Pro Glu Leu Pro Leu Phe Leu Lys Ala Tyr Trp His  
 290 295 300  
 His Gln Asn Leu Cys Gly Lys Ser Tyr Tyr Arg Arg Val Gly Arg Asp  
 305 310 315 320  
 Ala Gly Leu Ala Ala Ala Glu Arg Ser Ala  
 325 330

<210> 115  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 115  
 atgaaccaa tcattaaagc ggcggcagtt caatgtagcc ctgtgttgta tagccaagcg 60  
 ggtacagtca agaaaatctg tgacacgatt ttggagttgg ggcagcaagg tgtgcaattt 120  
 gccgtatttc ctgaaactgt tgtgccttat tacccttatt tttcttttgt gcaaccaccg 180  
 tttgccatgg gtaaagaaca tttaaagcta ttgcatgaat cggttgtcgt gccatcggca 240  
 gcaacaactt taattggaca ggcatgcaaa gaagcgaaca tgggtggttc tattggtatt 300  
 aatgagcgtg caggcggcac gatttataac gctcaattgt tgtttgatgc ggatggttcg 360  
 attattcagc atcgccgtaa aattacccca acgtatcatg aacgtatggt gtgggggcaa 420  
 ggcgatggca gtggttttacg tgcgatagat tctgctgtag gacgtattgg gtcgctggca 480  
 tgttgggagc attacaaccc tttggctcgg tttgctttga tggcggatgg tgagcaaatt 540  
 catgcggcga tgtttccggg atcactcgtg gggcagattt ttgcagatca gatcagtgcc 600  
 accattcagc accatgcttt agagtcgggc tgttttgtgg tgaatgccac agcatggctt 660  
 gacccagagc aacaacaaca aattatgcaa gatacaggct gtgaactcgg tccaatttcg 720  
 gggggatggt ttacggccat cgtttctcca gaaggcaa ttttgtctga accgatcaca 780  
 caaggcgaag gttatgtgat tgccgattta gacttttctt taatcgaaaa acgtaaacgg 840  
 atgatggatt ctgttgggca ttatagtcgt ccagaattac tcagtttggt gattgatcgt 900  
 cgtcctacct cagttttgca tgagttaaaa ctagagaatc catcgaataa cagcatcgaa 960  
 aaagtgtctg aatttgccga ggtacacgca tag 993

<210> 116  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 116  
 Met Asn Gln Ile Ile Lys Ala Ala Ala Val Gln Cys Ser Pro Val Leu  
 1 5 10 15  
 Tyr Ser Gln Ala Gly Thr Val Lys Lys Ile Cys Asp Thr Ile Leu Glu  
 20 25 30  
 Leu Gly Gln Gln Gly Val Gln Phe Ala Val Phe Pro Glu Thr Val Val  
 35 40 45  
 Pro Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Pro Pro Phe Ala Met Gly  
 50 55 60  
 Lys Glu His Leu Lys Leu Leu His Glu Ser Val Val Val Pro Ser Ala  
 65 70 75 80  
 Ala Thr Thr Leu Ile Gly Gln Ala Cys Lys Glu Ala Asn Met Val Val  
 85 90 95



Ser Ile Gly Ile Asn Glu Arg Ala Gly Gly Thr Ile Tyr Asn Ala Gln  
 100 105 110  
 Leu Leu Phe Asp Ala Asp Gly Ser Ile Ile Gln His Arg Arg Lys Ile  
 115 120 125  
 Thr Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp Gly Ser  
 130 135 140  
 Gly Leu Arg Ala Ile Asp Ser Ala Val Gly Arg Ile Gly Ser Leu Ala  
 145 150 155 160  
 Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Phe Ala Leu Met Ala Asp  
 165 170 175  
 Gly Glu Gln Ile His Ala Ala Met Phe Pro Gly Ser Leu Val Gly Gln  
 180 185 190  
 Ile Phe Ala Asp Gln Ile Ser Ala Thr Ile Gln His His Ala Leu Glu  
 195 200 205  
 Ser Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Pro Glu Gln  
 210 215 220  
 Gln Gln Gln Ile Met Gln Asp Thr Gly Cys Glu Leu Gly Pro Ile Ser  
 225 230 235 240  
 Gly Gly Cys Phe Thr Ala Ile Val Ser Pro Glu Gly Lys Phe Leu Ser  
 245 250 255  
 Glu Pro Ile Thr Gln Gly Glu Gly Tyr Val Ile Ala Asp Leu Asp Phe  
 260 265 270  
 Ser Leu Ile Glu Lys Arg Lys Arg Met Met Asp Ser Val Gly His Tyr  
 275 280 285  
 Ser Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Arg Pro Thr Ser  
 290 295 300  
 Val Leu His Glu Leu Lys Leu Glu Asn Pro Ser Asn Asn Ser Ile Glu  
 305 310 315 320  
 Lys Val Ser Glu Phe Ala Glu Val His Ala  
 325 330

<210> 117  
 <211> 957  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 117  
 atgactcaat ccaggataat tcgtgctgcg gcagcgcaga tcgctccgga tttgcagggt 60  
 ccaggtaaca cgatcgacaa agtttgccgc accatcagcg aggcggccgc aaaaggcgta 120  
 cagattattg ttttccctga aaccttggtg ccttattacc cttacttctc ttacatttca 180  
 ccgcccattc aacagggcaa agaacatttg cggctgtatg accatgcagt ggttgtgccc 240  
 ggctcggaaa ccgaggcaat ttcagctctt gccgcccac acaatatggt ggtggttttg 300  
 ggtgtgaacg agcgcgatca cggcacactt tacaacgcac aaattatttt caacagcgac 360  
 ggaaagattc tggtgaagcg ccgaaaaatt acaccaactt atcacgagcg gatggtgtgg 420  
 gggcagggtg acgcttcagg cttgaagggt gttgattccg cagtgggccc tgtgggtgca 480  
 ttggcctgtt gggaacacta caacccttg gctcgctatt gttgatggc ccagcacgaa 540  
 gaaattcact gtgcgcagtt tcccggttca ttggtggggc aagtttttgc cgaccaaattg 600  
 gaagtgacca ttcgtcacca cgcacttgag tcgggctgtt ttgtcatcaa cagcaccgct 660  
 tggctttctg aagaacaggt tcaaagtatt tcatccgaca gcgcattgca gaaagggtt 720  
 agaggcgggt gtttcacggc cattgtcagc cctgagggaa agctgttggc tgagccgctc 780  
 accgagggtg agggcatggt gatcgccgac ctgcacatgg cgttggttac gaaacgcaaa 840  
 cgcattgatg attcagtggg ccattatgcg cgccccgagt tggttgagttt gctggttcgg 900  
 gatgaggctt caagcccat gaaaaaaatt caggaggttc aacatgctga gtactga 957

<210> 118  
 <211> 318  
 <212> PRT  
 <213> Unknown



&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 118

```

Met Thr Gln Ser Arg Ile Ile Arg Ala Ala Ala Ala Gln Ile Ala Pro
 1           5           10           15
Asp Leu Gln Val Pro Gly Asn Thr Ile Asp Lys Val Cys Arg Thr Ile
      20           25           30
Ser Glu Ala Ala Ala Lys Gly Val Gln Ile Ile Val Phe Pro Glu Thr
      35           40           45
Leu Val Pro Tyr Tyr Pro Tyr Phe Ser Tyr Ile Ser Pro Pro Ile Gln
      50           55           60
Gln Gly Lys Glu His Leu Arg Leu Tyr Asp His Ala Val Val Val Pro
65           70           75           80
Gly Ser Glu Thr Glu Ala Ile Ser Ala Leu Ala Ala Gln His Asn Met
      85           90           95
Val Val Val Leu Gly Val Asn Glu Arg Asp His Gly Thr Leu Tyr Asn
      100          105          110
Ala Gln Ile Ile Phe Asn Ser Asp Gly Lys Ile Leu Leu Lys Arg Arg
      115          120          125
Lys Ile Thr Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp
      130          135          140
Ala Ser Gly Leu Lys Val Val Asp Ser Ala Val Gly Arg Val Gly Ala
145          150          155          160
Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Cys Leu Met
      165          170          175
Ala Gln His Glu Glu Ile His Cys Ala Gln Phe Pro Gly Ser Leu Val
      180          185          190
Gly Gln Val Phe Ala Asp Gln Met Glu Val Thr Ile Arg His His Ala
      195          200          205
Leu Glu Ser Gly Cys Phe Val Ile Asn Ser Thr Ala Trp Leu Ser Glu
      210          215          220
Glu Gln Val Gln Ser Ile Ser Ser Asp Ser Ala Leu Gln Lys Gly Leu
225          230          235          240
Arg Gly Gly Cys Phe Thr Ala Ile Val Ser Pro Glu Gly Lys Leu Leu
      245          250          255
Ala Glu Pro Leu Thr Glu Gly Glu Gly Met Val Ile Ala Asp Leu Asp
      260          265          270
Met Ala Leu Val Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly His
      275          280          285
Tyr Ala Arg Pro Glu Leu Leu Ser Leu Leu Val Arg Asp Glu Ala Ser
      290          295          300
Ser Pro Met Lys Lys Ile Gln Gly Val Gln His Ala Glu Tyr
305          310          315

```

&lt;210&gt; 119

&lt;211&gt; 984

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 119

```

atggatacac tcaaagttgg attggttcag atggccccc tctggttgaa ccgggataaa      60
accctgatca aagttgagga atacatgcag aaagcaggca aacagggctg caacctggta      120
gcttttggtg aagcgctggt tcccggtac cccttctggg tggaacgcac agagggcgcc      180
agattcaatt ccaaagtcca gaaagaactc ttgacacatt accttgatca ggcggtgcag      240
atcgaagccg gccaccttga tcctctccag gcattagccc aacaatacaa gatggctgtg      300
tacgtgggga cgattgaacg cccgcctgag cggagcggcc acagcctgta ctgctcccta      360

```



```

atatttatag acccagaagg cgagatcggc tcggttcacc gcaagttgat gcccacccat 420
gaggaacgcc tggctctggc aactggcgat gggcacggcc tgcgaacaca ttctctgggc 480
gcctttaccg ttggcggact caactgctgg gaaaactgga tgccgctctc ccgcacagct 540
ctttatgcca tgggagagga tcttcatgtt gctgcctggc ccgggagtca gcgcaatact 600
tatgatataa ccaaattcat tgccaaggaa tctcgtctct atgtgatctc cgtatccggg 660
atgatgaaaa aagaaaatat cctctctgaa attccccaca gccaatgat gctggaaaat 720
agcgaggata ttatggctga tggcggatcc tgtctggctg gaccagatgg agaatggatc 780
atcgagccca tcgtcggaga ggaaaccctg gtaactgctg aactatcaca tcagcgggtc 840
agagaagaaa gacagaattt cgaccaaca ggctactaca gtcggcctga tgtgaccgcg 900
ctggtagtcg accgcaggcg ccagcagatc ctggagatca ccccgacga aaaaggaaga 960
tcggatgaaa atcaatccct ttaa 984

```

&lt;210&gt; 120

&lt;211&gt; 327

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 120

```

Met Asp Thr Leu Lys Val Gly Leu Val Gln Met Ala Pro Ile Trp Leu
1          5          10          15
Asn Arg Asp Lys Thr Leu Ile Lys Val Glu Glu Tyr Met Gln Lys Ala
20          25          30
Gly Lys Gln Gly Cys Asn Leu Val Ala Phe Gly Glu Ala Leu Val Pro
35          40          45
Gly Tyr Pro Phe Trp Val Glu Arg Thr Glu Gly Ala Arg Phe Asn Ser
50          55          60
Lys Val Gln Lys Glu Leu Phe Ala His Tyr Leu Asp Gln Ala Val Gln
65          70          75          80
Ile Glu Ala Gly His Leu Asp Pro Leu Gln Ala Leu Ala Gln Gln Tyr
85          90          95
Lys Met Ala Val Tyr Val Gly Thr Ile Glu Arg Pro Pro Glu Arg Ser
100         105         110
Gly His Ser Leu Tyr Cys Ser Leu Ile Phe Ile Asp Pro Glu Gly Glu
115         120         125
Ile Gly Ser Val His Arg Lys Leu Met Pro Thr His Glu Glu Arg Leu
130         135         140
Val Trp Ser Thr Gly Asp Gly His Gly Leu Arg Thr His Ser Leu Gly
145         150         155         160
Ala Phe Thr Val Gly Gly Leu Asn Cys Trp Glu Asn Trp Met Pro Leu
165         170         175
Ser Arg Thr Ala Leu Tyr Ala Met Gly Glu Asp Leu His Val Ala Ala
180         185         190
Trp Pro Gly Ser Gln Arg Asn Thr Tyr Asp Ile Thr Lys Phe Ile Ala
195         200         205
Lys Glu Ser Arg Ser Tyr Val Ile Ser Val Ser Gly Met Met Lys Lys
210         215         220
Glu Asn Ile Leu Ser Glu Ile Pro His Ser Gln Leu Met Leu Glu Asn
225         230         235         240
Ser Glu Asp Ile Met Ala Asp Gly Gly Ser Cys Leu Ala Gly Pro Asp
245         250         255
Gly Glu Trp Ile Ile Glu Pro Ile Val Gly Glu Glu Thr Leu Val Thr
260         265         270
Ala Glu Leu Ser His Gln Arg Val Arg Glu Glu Arg Gln Asn Phe Asp
275         280         285
Pro Thr Gly His Tyr Ser Arg Pro Asp Val Thr Arg Leu Val Val Asp
290         295         300
Arg Arg Arg Gln Gln Ile Leu Glu Ile Thr Pro Asp Glu Lys Gly Arg
305         310         315         320

```



WO 03/097810

PCT/US03/15712

Ser Asp Glu Asn Gln Ser Leu  
325

<210> 121  
<211> 1158  
<212> DNA  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 121  
atgagcaaaa aagttctagg cggcagagaa aaagtaaaag ttgcagtagt tcaggctgcg 60  
cccgttttca tggacaagga gaagacgatt gaaaaggctt gcaagctaataaaaagaagcg 120  
gggagaaatc gagccgagct catagcgctt tcagagtcattcatccccgt ctatcctgca 180  
tactataaccg tcggctatga aaccccttct caagaatgga gagattacgt gattgcgcta 240  
caggataact ccgtgctgat tccgagcgag gataccgagg tactcggaca ggctgcaaag 300  
gaggcagggg cttatgcagt aataggatgc agcgagatgg acgaccgtcc gggaagccga 360  
acagtttaca acacgctcct cttcatcggc aaagacggca aggtcatggg aaggcataga 420  
aaactcaaac ccacgttcac ggagagaata tactggggag agggagatgc tggagacata 480  
aagggttttg ataccgagat cggcaggatc ggaggcctcg tatgctggga gaaccatatg 540  
actctagtca gggccgcat gatcacaggg ggagaggagt ttcataatcgc ggtctggccg 600  
ggaaactgga aggggtgcga aaacaagctt ctccaagcag ataataagccc aggagggcgc 660  
ctctgcaacc ttcaatctct cattaaagta cagcctttg aggccggggc gtttgtgctg 720  
agcgcttgcg gctttttgac gccagaggat ttcccggaaa ggtggcatta tataagggat 780  
ggtaaccata ttaactgcga ctgggcactg ggcggaagct caatcgtcaa tcccgccggc 840  
cgttatctcg tcgagcctaa ctttgagaag gatgcaatcc tctatgcgga ttgttatgca 900  
aaccagataa aagcagtaaa agcgggtttt gattcccttg gccactattc ccgctgggat 960  
attgcccaac tggcgataag gcaggaagcc tggaatccag aggtttcttt gatcgattcc 1020  
tcttcgactg aagttgagct tccggcagac gagcttcgaa ggatttcgga gaagtttgaa 1080  
gtaactgcgg ataagttgga atctttgctt gaggaaattg gaaagattaa aaagcccagg 1140  
aaacaagccg gttcctaa 1158

<210> 122  
<211> 385  
<212> PRT  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 122  
Met Ser Lys Lys Val Leu Gly Gly Arg Glu Lys Val Lys Val Ala Val  
1 5 10 15  
Val Gln Ala Ala Pro Val Phe Met Asp Lys Glu Lys Thr Ile Glu Lys  
20 25 30  
Ala Cys Lys Leu Ile Lys Glu Ala Gly Arg Asn Arg Ala Glu Leu Ile  
35 40 45  
Ala Phe Ser Glu Ser Phe Ile Pro Val Tyr Pro Ala Tyr Tyr Thr Val  
50 55 60  
Gly Tyr Glu Thr Pro Ser Gln Glu Trp Arg Asp Tyr Val Ile Ala Leu  
65 70 75 80  
Gln Asp Asn Ser Val Leu Ile Pro Ser Glu Asp Thr Glu Val Leu Gly  
85 90 95  
Gln Ala Ala Lys Glu Ala Gly Ala Tyr Ala Val Ile Gly Cys Ser Glu  
100 105 110  
Met Asp Asp Arg Pro Gly Ser Arg Thr Val Tyr Asn Thr Leu Leu Phe  
115 120 125  
Ile Gly Lys Asp Gly Lys Val Met Gly Arg His Arg Lys Leu Lys Pro  
130 135 140  
Thr Phe Thr Glu Arg Ile Tyr Trp Gly Glu Gly Asp Ala Gly Asp Ile



145 150 155 160  
 Lys Val Phe Asp Thr Glu Ile Gly Arg Ile Gly Gly Leu Val Cys Trp  
 165 170 175  
 Glu Asn His Met Thr Leu Val Arg Ala Ala Met Ile His Arg Gly Glu  
 180 185 190  
 Glu Phe His Ile Ala Val Trp Pro Gly Asn Trp Lys Gly Ala Glu Asn  
 195 200 205  
 Lys Leu Leu Gln Ala Asp Asn Ser Pro Gly Gly Ala Leu Cys Asn Leu  
 210 215 220  
 Gln Ser Leu Ile Lys Val His Ala Phe Glu Ala Gly Ala Phe Val Leu  
 225 230 235 240  
 Ser Ala Cys Gly Phe Leu Thr Pro Glu Asp Phe Pro Glu Arg Trp His  
 245 250 255  
 Tyr Ile Arg Asp Gly Asn His Ile Asn Cys Asp Trp Ala Leu Gly Gly  
 260 265 270  
 Ser Ser Ile Val Asn Pro Ala Gly Arg Tyr Leu Val Glu Pro Asn Phe  
 275 280 285  
 Glu Lys Asp Ala Ile Leu Tyr Ala Asp Cys Tyr Ala Asn Gln Ile Lys  
 290 295 300  
 Ala Val Lys Ala Val Phe Asp Ser Leu Gly His Tyr Ser Arg Trp Asp  
 305 310 315 320  
 Ile Ala Gln Leu Ala Ile Arg Gln Glu Ala Trp Asn Pro Glu Val Ser  
 325 330 335  
 Leu Ile Asp Ser Ser Ser Thr Glu Val Glu Leu Pro Ala Asp Glu Leu  
 340 345 350  
 Arg Arg Ile Ser Glu Lys Phe Glu Val Thr Ala Asp Lys Leu Glu Ser  
 355 360 365  
 Leu Leu Glu Glu Ile Gly Lys Ile Lys Lys Pro Arg Lys Gln Ala Gly  
 370 375 380  
 Ser  
 385

<210> 123  
 <211> 990  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 123  
 atgtcaactt tcaagatcgc taccgtgcag agtgcaccag tatattatgga ccgcgaagct 60  
 accattgaca agacttgcca gctgatcgcc gaagcagcac aagatgacga cgttcgccta 120  
 gtggtcttcc ccgaggcctt tatccccacc tatccggact gggatatggcg tatccctccc 180  
 ggacagcacc agatgcttgc cgacctgtac ggggagttgc tcgagcagtc ggtgacgata 240  
 cccagtctgg ctaccgagcg gctctgtcag gctgcaaaga aagcgggcgt ttatgtagct 300  
 gtgggcctta acgaacgcaa tacagaggcc agcaacgcta ccctgtacaa caccctgctc 360  
 tacattgacg ccgagggcaa cttgctaggt aagcaccgaa agctggtacc gaccgctccc 420  
 gaacgcatgg tctgggcaca gggagatggc agtacccttg aggtctacga gacctccttc 480  
 ggaaaactca gcggactaat ctggtgggag aactacatgc ctctcgctcg ttatgccctg 540  
 tatgcctggg gagtacagct ctatttggct cctacttggg atcgaggcga gccctggctt 600  
 tccactctgc ggcacattgc caaggaagga cgagtatacg tggtcggctg ctctatcgcc 660  
 ttacgtaagg aagacatccc cgaccgattc gaattcaagg cgaagtacta cgcagaggca 720  
 ggagagtggg taaacaaagg tgacagcgtc atcgtcggtc ccgatggcga gctcatcgcc 780  
 gggcctctac ataaggaaca ggggatactc tatgctgagc tggacacaag gcagatgcac 840  
 gccccaagt ggaacctgga tgtagccgga cactacgcgc gcccgacgt gtttcggctg 900  
 accgtgagca aggatggcca tccgatgctc ggcgttgccc aagggcccaa gcatgagccg 960  
 caagataaga ccgaagtatt agagggctag 990

<210> 124  
 <211> 329



WO 03/097810

PCT/US03/15712

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 124

```

Met Ser Thr Phe Lys Ile Ala Thr Val Gln Ser Ala Pro Val Phe Met
 1          5          10          15
Asp Arg Glu Ala Thr Ile Asp Lys Thr Cys Glu Leu Ile Ala Glu Ala
      20          25          30
Ala Gln Asp Asp Asp Val Arg Leu Val Val Phe Pro Glu Ala Phe Ile
      35          40          45
Pro Thr Tyr Pro Asp Trp Val Trp Arg Ile Pro Pro Gly Gln His Gln
      50          55          60
Met Leu Ala Asp Leu Tyr Gly Glu Leu Leu Glu Gln Ser Val Thr Ile
65          70          75          80
Pro Ser Leu Ala Thr Glu Arg Leu Cys Gln Ala Ala Lys Lys Ala Gly
      85          90          95
Val Tyr Val Ala Val Gly Leu Asn Glu Arg Asn Thr Glu Ala Ser Asn
      100          105          110
Ala Thr Leu Tyr Asn Thr Leu Leu Tyr Ile Asp Ala Glu Gly Asn Leu
      115          120          125
Leu Gly Lys His Arg Lys Leu Val Pro Thr Ala Pro Glu Arg Met Val
      130          135          140
Trp Ala Gln Gly Asp Gly Ser Thr Leu Glu Val Tyr Glu Thr Ser Phe
145          150          155          160
Gly Lys Leu Ser Gly Leu Ile Cys Trp Glu Asn Tyr Met Pro Leu Ala
      165          170          175
Arg Tyr Ala Leu Tyr Ala Trp Gly Val Gln Leu Tyr Leu Ala Pro Thr
      180          185          190
Trp Asp Arg Gly Glu Pro Trp Leu Ser Thr Leu Arg His Ile Ala Lys
      195          200          205
Glu Gly Arg Val Tyr Val Val Gly Cys Ser Ile Ala Leu Arg Lys Glu
      210          215          220
Asp Ile Pro Asp Arg Phe Glu Phe Lys Ala Lys Tyr Tyr Ala Glu Ala
225          230          235          240
Gly Glu Trp Ile Asn Lys Gly Asp Ser Val Ile Val Gly Pro Asp Gly
      245          250          255
Glu Leu Ile Ala Gly Pro Leu His Lys Glu Gln Gly Ile Leu Tyr Ala
      260          265          270
Glu Leu Asp Thr Arg Gln Met His Ala Pro Lys Trp Asn Leu Asp Val
      275          280          285
Ala Gly His Tyr Ala Arg Pro Asp Val Phe Arg Leu Thr Val Ser Lys
      290          295          300
Asp Gly His Pro Met Leu Gly Val Ala Gln Gly Pro Lys His Glu Pro
305          310          315          320
Gln Asp Lys Thr Glu Val Leu Glu Gly
      325

```

&lt;210&gt; 125

&lt;211&gt; 1050

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 125

```

atgacaactg taaaaaagac ggtacgcgca gcagcgatcc agatcgcacc tgacctcgac      60
agtgcaggcg gtacgctgga caaggtttgc acggccattc aaaaggcggc ggcacaaggc      120

```



```

gcgagctgg tggtttttcc cgaaaccttc ttgccctact atccttactt ttcattcgtg 180
cgcccgccct tcgcatccgg cccggaacac ttgctgctat atgaacgcgc agtggcggtg 240
ccaggcccgg tgaccgatgc cgtctctgcc gtcgcgcgca gccacggcgt ggtggtggta 300
ctcggcgtca atgaacgcga ccatggcacg ctgtacaaca cccaactggt gttcgacgcg 360
aatggcgaac tgggtgtgaa acgcagaaaa atcacgccga cttatcacga gcggatgata 420
tggtgtcaag gcgacggcag cggactcaaa gtagtgcaaa cggcggtcgg ccggctaggc 480
gcgctagcct gttgggaaca ctacaaccca ctggcccgtt atgcattgat ggcgcaacac 540
gaagaaatcc attgcgcca gtttcccggg tccatggtcg ggcaaatatt cggcgaccag 600
atggaagtga cgatacgcca tcacgctctc gagtcggctt gcttcgtggt gaatgccaca 660
ggctggctga ccgatgcgca aatcacatcg atcacgccgg accccgcgct acaaaaggca 720
ttacgtggcg gttgctgcac cgccatcgtc tcgccggaag gtgtgctcct ggcagagccg 780
ctacgcagcg gcgaaggcat ggtgatcgcc gatctcgata tggcactcat caccaaacgc 840
aaacggatga tggattcggg cggccactat gcgcggcccg aattgttaag cctgcttgtc 900
gacgaccggc gcaaggtacc ggtatccgcg ctatttgccg acagcaacc tgccaacggg 960
cacacagttt tcaccccatc cgacatacca acccttggga gcgcacatca tgcaaacagt 1020
taccaaaccg aaccagcaac tgatcactga 1050

```

&lt;210&gt; 126

&lt;211&gt; 349

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 126

```

Met Thr Thr Val Lys Lys Thr Val Arg Ala Ala Ala Ile Gln Ile Ala
1      5      10      15
Pro Asp Leu Asp Ser Ala Gly Gly Thr Leu Asp Lys Val Cys Thr Ala
20     25     30
Ile Gln Lys Ala Ala Ala Gln Gly Ala Glu Leu Val Val Phe Pro Glu
35     40     45
Thr Phe Leu Pro Tyr Tyr Pro Tyr Phe Ser Phe Val Arg Pro Pro Phe
50     55     60
Ala Ser Gly Pro Glu His Leu Leu Leu Tyr Glu Arg Ala Val Ala Val
65     70     75     80
Pro Gly Pro Val Thr Asp Ala Val Ser Ala Val Ala Arg Ser His Gly
85     90     95
Val Val Val Val Leu Gly Val Asn Glu Arg Asp His Gly Thr Leu Tyr
100    105    110
Asn Thr Gln Leu Val Phe Asp Ala Asn Gly Glu Leu Val Leu Lys Arg
115    120    125
Arg Lys Ile Thr Pro Thr Tyr His Glu Arg Met Ile Trp Gly Gln Gly
130    135    140
Asp Gly Ser Gly Leu Lys Val Val Gln Thr Ala Val Gly Arg Leu Gly
145    150    155    160
Ala Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu
165    170    175
Met Ala Gln His Glu Glu Ile His Cys Ala Gln Phe Pro Gly Ser Met
180    185    190
Val Gly Gln Ile Phe Ala Asp Gln Met Glu Val Thr Ile Arg His His
195    200    205
Ala Leu Glu Ser Ala Cys Phe Val Val Asn Ala Thr Gly Trp Leu Thr
210    215    220
Asp Ala Gln Ile Thr Ser Ile Thr Pro Asp Pro Ala Leu Gln Lys Ala
225    230    235    240
Leu Arg Gly Gly Cys Cys Thr Ala Ile Val Ser Pro Glu Gly Val Leu
245    250    255
Leu Ala Glu Pro Leu Arg Ser Gly Glu Gly Met Val Ile Ala Asp Leu
260    265    270
Asp Met Ala Leu Ile Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly

```



WO 03/097810

PCT/US03/15712

275 280 285  
 His Tyr Ala Arg Pro Glu Leu Leu Ser Leu Leu Val Asp Asp Arg Arg  
 290 295 300  
 Lys Val Pro Val Ser Ala Leu Phe Ala Asp Ser Asn Pro Ala Asn Gly  
 305 310 315 320  
 His Thr Val Phe Thr Pro Ser Asp Ile Pro Thr Leu Gly Ser Ala His  
 325 330 335  
 His Ala Asn Ser Tyr Gln Thr Glu Pro Ala Thr Asp His  
 340 345

<210> 127  
 <211> 1005  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 127  
 atgatagcac ggaagacaat aagggcggcg gcggtgcaga tagcgctgt gatggaagat 60  
 cggaaggcga cgaccgacaa ggtgtgcgcc tacattcagg aagcaggcga gaatggagcc 120  
 gaaattgtgg tgtttcctga aaccttcatt cccaattatc cctatttctc ttttgtaaaa 180  
 cctcccgtgt tggcaggtaa ggatcacctt accttgtatg accaagcggg ggaaatccct 240  
 agccctacta ccgaccaagt ggggtctatg gccaaaaaat ggggaatcgt agtgggtgtg 300  
 ggcgtgaacg aaagaagcca cggcactttg tacaatgccc aaattgtctt tgacgctact 360  
 ggtgatattg tattggtgag acgcaaaatc acccctacct atcatgaacg gatgatctgg 420  
 ggacagggag atggcagtgg attaaaagca gtagacacag ctgtgggaag agtgggcgct 480  
 ttggcgtgtt gggaacacta taatccactt gcgcgctacg cccttatggt agaccatgag 540  
 gaaattcatt gcagccaatt ccctggctct atggtcggcc ccattttcgg tgaccagata 600  
 gaagtgaacg ttgcgccacca tgcgttgga tccgggtgtt ttgtcatcaa ttccacaggt 660  
 tggctgtttg aagagcaaat ccaagccatc accgatgatc cgaaactgca caaagcattg 720  
 aaagacggct gtatgaccgc cattatttct cccgaaggcg tgcatttgac caaacctta 780  
 acagaaggcg aaggcatcat ctacgcctat ctggacatga aactcataga caagcggaaa 840  
 cggatgatgg actcggtagg acactatgca cgtccagagt tgctctcttt gcatatcaac 900  
 aatgcagagc aaaaaccagc cgtttacacc tctcctctta ccaaaacgga aaccaagaa 960  
 gacgtaaaaa gctatgatcg caacaaagaa cagcttatcg tctga 1005

<210> 128  
 <211> 334  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 128  
 Met Ile Ala Arg Lys Thr Ile Arg Ala Ala Ala Val Gln Ile Ala Pro  
 1 5 10 15  
 Val Met Glu Asp Arg Lys Ala Thr Thr Asp Lys Val Cys Ala Tyr Ile  
 20 25 30  
 Gln Glu Ala Gly Glu Asn Gly Ala Glu Ile Val Val Phe Pro Glu Thr  
 35 40 45  
 Phe Ile Pro Asn Tyr Pro Tyr Phe Ser Phe Val Lys Pro Pro Val Leu  
 50 55 60  
 Ala Gly Lys Asp His Leu Thr Leu Tyr Asp Gln Ala Val Glu Ile Pro  
 65 70 75 80  
 Ser Pro Thr Thr Asp Gln Val Gly Ser Met Ala Lys Lys Trp Gly Ile  
 85 90 95  
 Val Val Val Leu Gly Val Asn Glu Arg Ser His Gly Thr Leu Tyr Asn  
 100 105 110  
 Ala Gln Ile Val Phe Asp Ala Thr Gly Asp Ile Val Leu Val Arg Arg







<223> Obtained from an environmental sample

<400> 130

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Glu | Lys | Arg | Ile | Ile | Lys | Ala | Ala | Ala | Val | Gln | Ile | Thr | Pro |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     | 15  |     |     |
| Asp | Phe | Glu | Ser | His | Asp | Gly | Thr | Val | Lys | Lys | Val | Cys | Asn | Val | Ile |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Asp | Glu | Ala | Gly | Ala | Lys | Gly | Val | Gln | Ile | Ile | Val | Phe | Pro | Glu | Thr |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Phe | Ile | Pro | Tyr | Tyr | Pro | Tyr | Phe | Ser | Phe | Ile | Thr | Pro | Pro | Val | Thr |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Ala | Gly | Ala | Glu | His | Leu | Arg | Leu | Tyr | Glu | Lys | Ser | Val | Val | Ile | Pro |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Gly | Pro | Val | Thr | Gln | Ala | Ile | Ser | Glu | Arg | Ala | Arg | Met | Asn | Asn | Met |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Val | Val | Val | Leu | Gly | Val | Asn | Glu | Arg | Asp | Asn | Gly | Ser | Leu | Tyr | Asn |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Thr | Gln | Ile | Ile | Phe | Asp | Ala | Thr | Gly | Glu | Met | Leu | Leu | Lys | Arg | Arg |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     |     | 125 |     |     |
| Lys | Ile | Thr | Pro | Thr | Tyr | His | Glu | Arg | Met | Ile | Trp | Gly | Gln | Gly | Asp |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Ala | Ser | Gly | Leu | Lys | Val | Val | Asp | Thr | Ala | Ile | Gly | Arg | Val | Gly | Ala |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Leu | Ala | Cys | Trp | Glu | His | Tyr | Asn | Pro | Leu | Ala | Arg | Tyr | Ser | Leu | Met |
|     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Thr | Gln | His | Glu | Glu | Ile | His | Cys | Ala | Gln | Phe | Pro | Gly | Ser | Met | Val |
|     |     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |     |
| Gly | Gln | Ile | Phe | Ala | Asp | Gln | Met | Asp | Val | Thr | Ile | Arg | His | His | Ala |
|     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Leu | Glu | Ser | Gly | Cys | Phe | Val | Ile | Asn | Ser | Thr | Gly | Trp | Leu | Thr | Asp |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Asp | Gln | Ile | Lys | Ser | Ile | Thr | Asp | Asp | Pro | Lys | Met | Gln | Lys | Ala | Leu |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Arg | Gly | Gly | Cys | Asn | Thr | Ala | Ile | Ile | Ser | Pro | Glu | Gly | Asn | His | Leu |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Thr | Glu | Pro | Leu | Arg | Glu | Gly | Glu | Gly | Met | Val | Ile | Ala | Asp | Leu | Asp |
|     |     | 260 |     |     |     |     | 265 |     |     |     |     |     | 270 |     |     |
| Met | Ala | Leu | Ile | Thr | Lys | Arg | Lys | Arg | Met | Met | Asp | Ser | Val | Gly | His |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Tyr | Ala | Arg | Pro | Glu | Leu | Leu | Ser | Leu | Ala | Ile | Asn | Asp | Ala | Pro | Ala |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Thr | Pro | Ser | Phe | Gln | Met | Asn | Glu | His | Arg | Leu | Lys | Ser | Val | Gln | Leu |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Pro | Ile | Ala | Glu | Glu | Leu | Lys | Asn | Asp | Asn | Lys | Leu | Ser | Ser | Gly | Gln |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |

<210> 131

<211> 1011

<212> DNA

<213> Unknown

<220>

<223> Obtained from an environmental sample

<400> 131

|            |            |            |            |            |             |     |
|------------|------------|------------|------------|------------|-------------|-----|
| atgtcagaaa | agcgaattat | taaagcggct | gcagttcaaa | tcacaccaga | ttttgaatcg  | 60  |
| catgatggaa | ccgtaaagaa | ggtttgtaat | gtaattgatg | aagcgggtgc | ttaaagggtga | 120 |
| cagatcattg | tattccctga | aacctttatt | ccatattacc | catatttttc | tttcatcact  | 180 |
| ccaccagtga | ctgctggcgc | ggagcatttg | cggctctatg | aaaaaagtgt | cgtgatacct  | 240 |
| ggtcccgtta | ctcaagacat | ttccgaacgt | gcacgcatga | ataatatggg | tggtgtactt  | 300 |
| ggtgtaaatg | agcgtgataa | cggcagtcta | tataacaccc | agattatttt | tgatgctacc  | 360 |



```

ggtgagatgc ttctgaagag aagaaaaatc acacctacct atcatgagcg catgatttgg 420
gggcaaggag atgcttcagg cctgaagggtc gtcgatacgg ctattgggcg agtcggagca 480
ttggcatgct gggagcacta taaccctttg gctagataca gcctcatgac acagcatgaa 540
gaaattcact gtgctcaatt tccaggctcc atgggttggtc agatcttcgc agatcaaagt 600
gatgtcacga ttcgtcatca tgccttggag tcagggttgc tgcgtcatcaa ctccactggc 660
tggttaactg atgatcagat caaatctatc accgacgatc ccaaaatgca gaaagcttta 720
agaggtggtt gcaacacggc cattatttct ccagaaggga atcatttaac cgagcctttg 780
cgagaagggtg aaggcatggt gattgctgat cttgatattg cactcatcac caaacgaaaa 840
agaatgatgg actcagttgg ccactacgcc agaccagaac tggtgagctt agcgatcaat 900
gatgctccgg ctactccttc attccagatg aacgaacatc gtcttaaatac agtgcaatta 960
cctatcgcag aggagcttaa aatgacaac aagcttagca gtggacagta a 1011

```

&lt;210&gt; 132

&lt;211&gt; 336

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 132

```

Met Ser Glu Lys Arg Ile Ile Lys Ala Ala Ala Val Gln Ile Thr Pro
1          5          10          15
Asp Phe Glu Ser His Asp Gly Thr Val Lys Lys Val Cys Asn Val Ile
20          25          30
Asp Glu Ala Gly Ala Lys Gly Val Gln Ile Ile Val Phe Pro Glu Thr
35          40          45
Phe Ile Pro Tyr Tyr Pro Tyr Phe Ser Phe Ile Thr Pro Pro Val Thr
50          55          60
Ala Gly Ala Glu His Leu Arg Leu Tyr Glu Lys Ser Val Val Ile Pro
65          70          75          80
Gly Pro Val Thr Gln Asp Ile Ser Glu Arg Ala Arg Met Asn Asn Met
85          90          95
Val Val Val Leu Gly Val Asn Glu Arg Asp Asn Gly Ser Leu Tyr Asn
100         105         110
Thr Gln Ile Ile Phe Asp Ala Thr Gly Glu Met Leu Leu Lys Arg Arg
115         120         125
Lys Ile Thr Pro Thr Tyr His Glu Arg Met Ile Trp Gly Gln Gly Asp
130         135         140
Ala Ser Gly Leu Lys Val Val Asp Thr Ala Ile Gly Arg Val Gly Ala
145         150         155         160
Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ser Leu Met
165         170         175
Thr Gln His Glu Glu Ile His Cys Ala Gln Phe Pro Gly Ser Met Val
180         185         190
Gly Gln Ile Phe Ala Asp Gln Met Asp Val Thr Ile Arg His His Ala
195         200         205
Leu Glu Ser Gly Cys Phe Val Ile Asn Ser Thr Gly Trp Leu Thr Asp
210         215         220
Asp Gln Ile Lys Ser Ile Thr Asp Asp Pro Lys Met Gln Lys Ala Leu
225         230         235         240
Arg Gly Gly Cys Asn Thr Ala Ile Ile Ser Pro Glu Gly Asn His Leu
245         250         255
Thr Glu Pro Leu Arg Glu Gly Glu Gly Met Val Ile Ala Asp Leu Asp
260         265         270
Met Ala Leu Ile Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly His
275         280         285
Tyr Ala Arg Pro Glu Leu Leu Ser Leu Ala Ile Asn Asp Ala Pro Ala
290         295         300
Thr Pro Ser Phe Gln Met Asn Glu His Arg Leu Lys Ser Val Gln Leu
305         310         315         320

```



Pro Ile Ala Glu Glu Leu Lys Asn Asp Asn Lys Leu Ser Ser Gly Gln  
 325 330 335

<210> 133  
 <211> 1026  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 133  
 atgtcgacca agcggatcgt acgcgccgct gccgttcagc tggcaccgga tctggagcgg 60  
 ccggagggca cactggagaa ggtttgcgcg gccatcgaca aggcggcggg ggacgggtgtg 120  
 cagctcatcg tcttccccga gaccttcgta ccgtactacc cgtacttctc tttcgtgcgt 180  
 gcgccgggtcg cgatgggtgc cgagcacatg cggttatacg agcgcgcggt agcggtgccc 240  
 ggtccagtaa cggccaccgt ggcggagcgg gcaaaagcgc acgcgatggg cgtcgtgctg 300  
 ggtgtaaacg agcgcgatca cggctcactg tataacgcgc aactgatctt cgacgagacc 360  
 ggccgtctcg tcctcaaacg ccgcaagatc actccgacct atcacgagcg catggtgtgg 420  
 gggcagggcg acggcagcgg ccttaagggt gtagacaccg gtatcggcag gatcggagcc 480  
 ctcgcctgct gggagcacta caaccgcgtc gcgcgctatg cgctcatggc gcagcacgaa 540  
 gagattcatt gcgcgcagtt tccgggctcg atggtggggc cgatcttcgc ggatcagatc 600  
 gaggtcacga tccgccatca cgcgctggag tcgggctgct tcgtcgtcaa tgcgaccggc 660  
 tggctgacac ccgaacagat cgcgctcgatc acaccggacg cgggtctgca aaaggcaatc 720  
 agcggggggt gcaacaccgc gatcatctcg ccggagggcg tgcacctggc cccgccgttg 780  
 cgagaagggt agggcatggt cgtggccgac ctcgacatgg cgctcatcac caaacgcaaa 840  
 cgcgatgatg attcgggtggg tctactacgct cgcccggagt tgctcagcct gcgcatcgat 900  
 agccgcgccg cttcgccgat gtcgtcacia atggaaatac ccgggagctt gcatgaaatc 960  
 accagccacg atgtccagcc agcaactgat gaccgagctc cagtcctccg gcttgagggt 1020  
 ggctga 1026

<210> 134  
 <211> 341  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 134  
 Met Ser Thr Lys Arg Ile Val Arg Ala Ala Ala Val Gln Leu Ala Pro  
 1 5 10 15  
 Asp Leu Glu Arg Pro Glu Gly Thr Leu Glu Lys Val Cys Ala Ala Ile  
 20 25 30  
 Asp Lys Ala Ala Gly Asp Gly Val Gln Leu Ile Val Phe Pro Glu Thr  
 35 40 45  
 Phe Val Pro Tyr Tyr Pro Tyr Phe Ser Phe Val Arg Ala Pro Val Ala  
 50 55 60  
 Met Gly Ala Glu His Met Arg Leu Tyr Glu Arg Ala Val Ala Val Pro  
 65 70 75 80  
 Gly Pro Val Thr Ala Thr Val Ala Glu Arg Ala Lys Ala His Ala Met  
 85 90 95  
 Val Val Val Leu Gly Val Asn Glu Arg Asp His Gly Ser Leu Tyr Asn  
 100 105 110  
 Ala Gln Leu Ile Phe Asp Glu Thr Gly Arg Leu Val Leu Lys Arg Arg  
 115 120 125  
 Lys Ile Thr Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp  
 130 135 140  
 Gly Ser Gly Leu Lys Val Val Asp Thr Gly Ile Gly Arg Ile Gly Ala  
 145 150 155 160  
 Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |  |  |  |  |
| Ala | Gln | His | Glu | Glu | Ile | His | Cys | Ala | Gln | Phe | Pro | Gly | Ser | Met | Val |  |  |  |  |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |  |  |  |  |
| Gly | Pro | Ile | Phe | Ala | Asp | Gln | Ile | Glu | Val | Thr | Ile | Arg | His | His | Ala |  |  |  |  |
|     |     | 195 |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |  |  |  |  |
| Leu | Glu | Ser | Gly | Cys | Phe | Val | Val | Asn | Ala | Thr | Gly | Trp | Leu | Thr | Pro |  |  |  |  |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |  |  |  |
| Glu | Gln | Ile | Ala | Ser | Ile | Thr | Pro | Asp | Ala | Gly | Leu | Gln | Lys | Ala | Ile |  |  |  |  |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |  |  |  |
| Ser | Gly | Gly | Cys | Asn | Thr | Ala | Ile | Ile | Ser | Pro | Glu | Gly | Val | His | Leu |  |  |  |  |
|     |     |     | 245 |     |     |     |     | 250 |     |     |     |     |     | 255 |     |  |  |  |  |
| Ala | Pro | Pro | Leu | Arg | Glu | Gly | Glu | Gly | Met | Val | Val | Ala | Asp | Leu | Asp |  |  |  |  |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |  |  |  |  |
| Met | Ala | Leu | Ile | Thr | Lys | Arg | Lys | Arg | Met | Met | Asp | Ser | Val | Gly | His |  |  |  |  |
|     |     | 275 |     |     |     | 280 |     |     |     |     |     | 285 |     |     |     |  |  |  |  |
| Tyr | Ala | Arg | Pro | Glu | Leu | Leu | Ser | Leu | Arg | Ile | Asp | Ser | Arg | Ala | Ala |  |  |  |  |
|     | 290 |     |     |     | 295 |     |     |     |     |     | 300 |     |     |     |     |  |  |  |  |
| Ser | Pro | Met | Ser | Ser | Gln | Met | Glu | Ile | Pro | Gly | Ser | Leu | His | Glu | Ile |  |  |  |  |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |  |  |  |
| Thr | Ser | His | Asp | Val | Gln | Pro | Ala | Thr | Asp | Asp | Arg | Ala | Pro | Val | Leu |  |  |  |  |
|     |     |     | 325 |     |     |     |     | 330 |     |     |     |     |     | 335 |     |  |  |  |  |
| Arg | Leu | Glu | Val | Gly |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |
|     |     |     | 340 |     |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |

<210> 135  
 <211> 1011  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 135  
 atgtcagaca agcgaatcat taaagcggct gcagttcaaa tcactcctga ctttgactca 60  
 gcagatggaa ccgttaagaa agtgtgcaag gtaatcgatg aagcaggtgc aaagggagtt 120  
 caaattattg tattcccgga aaccttcatc ccttactacc catacttttc attcattaca 180  
 cctccagtca ctgctggcgc tgagcattta aagctttatg agaaaagtgt cgtgatacct 240  
 ggcccgggta cccaagcgat tgccgagcga gccagggtta atcagatggg tgtcgtgctt 300  
 ggtgtcaacg agcgagataa cggtagcctc tacaacacac aattgatctt tgataccaac 360  
 ggcgaactgc tacttaaaag aagaaaaatc acccctacct accatgaacg tatgatctgg 420  
 gggcaagggtg atgcatcagg tctcaaagta gttgaaacag agatcgcccg agtaggtgcc 480  
 ttggcttggt gggaacacta caaccactg gccagatatg cactcatgac acagcatgaa 540  
 gaaattcact gtgcgcaatt cccaggctct atggttggcc agatatttgc cgatcagatg 600  
 gatgtcacta tccgacatca cgccttagag tcaggctgct tcgtcatcaa cgccactggc 660  
 tggctcaccg acgcgcaa atccaatcgatt actgatgacc caaaaatgca aaaagcatta 720  
 cgtggcggct gcaacacagc catcatctcc cccgaagggg tgcacttaac agagccacta 780  
 cgtgaaggag aaggcatggg gattgccaat cttgatattg cactcatcac aaaacgaaaa 840  
 agaatgatgg attcggtagg ccattattca agaccagaat tattaagcct ggcaattaac 900  
 gacaaaccag caactacaac attttcaatg actgaggggc gtactcaaac agagccattt 960  
 cgaatcgagcaggagggttgaa aatgacgac aagcttagca ctggaaacta a 1011

<210> 136  
 <211> 336  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 136  
 Met Ser Asp Lys Arg Ile Ile Lys Ala Ala Ala Val Gln Ile Thr Pro



|                                                                 |                                                     |     |     |
|-----------------------------------------------------------------|-----------------------------------------------------|-----|-----|
| 1                                                               | 5                                                   | 10  | 15  |
| Asp Phe Asp                                                     | Ser Ala Asp Gly Thr Val Lys Lys Val Cys Lys Val Ile |     |     |
|                                                                 | 20                                                  | 25  | 30  |
| Asp Glu Ala                                                     | Gly Ala Lys Gly Val Gln Ile Ile Val Phe Pro Glu Thr |     |     |
|                                                                 | 35                                                  | 40  | 45  |
| Phe Ile Pro Tyr Tyr Pro Tyr Phe Ser Phe Ile Thr Pro Pro Val Thr |                                                     |     |     |
|                                                                 | 50                                                  | 55  | 60  |
| Ala Gly Ala Glu His Leu Lys Leu Tyr Glu Lys Ser Val Val Ile Pro |                                                     |     |     |
| 65                                                              | 70                                                  | 75  | 80  |
| Gly Pro Val Thr Gln Ala Ile Ala Glu Arg Ala Arg Val Asn Gln Met |                                                     |     |     |
|                                                                 | 85                                                  | 90  | 95  |
| Val Val Val Leu Gly Val Asn Glu Arg Asp Asn Gly Ser Leu Tyr Asn |                                                     |     |     |
|                                                                 | 100                                                 | 105 | 110 |
| Thr Gln Leu Ile Phe Asp Thr Asn Gly Glu Leu Leu Leu Lys Arg Arg |                                                     |     |     |
|                                                                 | 115                                                 | 120 | 125 |
| Lys Ile Thr Pro Thr Tyr His Glu Arg Met Ile Trp Gly Gln Gly Asp |                                                     |     |     |
|                                                                 | 130                                                 | 135 | 140 |
| Ala Ser Gly Leu Lys Val Val Glu Thr Glu Ile Ala Arg Val Gly Ala |                                                     |     |     |
| 145                                                             | 150                                                 | 155 | 160 |
| Leu Ala Cys Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met |                                                     |     |     |
|                                                                 | 165                                                 | 170 | 175 |
| Thr Gln His Glu Ile His Cys Ala Gln Phe Pro Gly Ser Met Val     |                                                     |     |     |
|                                                                 | 180                                                 | 185 | 190 |
| Gly Gln Ile Phe Ala Asp Gln Met Asp Val Thr Ile Arg His His Ala |                                                     |     |     |
|                                                                 | 195                                                 | 200 | 205 |
| Leu Glu Ser Gly Cys Phe Val Ile Asn Ala Thr Gly Trp Leu Thr Asp |                                                     |     |     |
|                                                                 | 210                                                 | 215 | 220 |
| Ala Gln Ile Gln Ser Ile Thr Asp Asp Pro Lys Met Gln Lys Ala Leu |                                                     |     |     |
| 225                                                             | 230                                                 | 235 | 240 |
| Arg Gly Gly Cys Asn Thr Ala Ile Ile Ser Pro Glu Gly Val His Leu |                                                     |     |     |
|                                                                 | 245                                                 | 250 | 255 |
| Thr Glu Pro Leu Arg Glu Gly Glu Gly Met Val Ile Ala Asn Leu Asp |                                                     |     |     |
|                                                                 | 260                                                 | 265 | 270 |
| Met Ala Leu Ile Thr Lys Arg Lys Arg Met Met Asp Ser Val Gly His |                                                     |     |     |
|                                                                 | 275                                                 | 280 | 285 |
| Tyr Ser Arg Pro Glu Leu Leu Ser Leu Ala Ile Asn Asp Lys Pro Ala |                                                     |     |     |
|                                                                 | 290                                                 | 295 | 300 |
| Thr Thr Thr Phe Ser Met Thr Glu Gly Arg Thr Gln Thr Glu Pro Phe |                                                     |     |     |
| 305                                                             | 310                                                 | 315 | 320 |
| Arg Ile Ala Glu Glu Leu Lys Asn Asp Asp Lys Leu Ser Thr Gly Asn |                                                     |     |     |
|                                                                 | 325                                                 | 330 | 335 |

&lt;210&gt; 137

&lt;211&gt; 978

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 137

|            |            |             |             |             |             |     |
|------------|------------|-------------|-------------|-------------|-------------|-----|
| atggctattg | tcaaggccgc | ggcgggtgcag | atcagtcgag  | tgctctacag  | tcgcgccggc  | 60  |
| acagtggaca | aggtcgctgc | gaagatccgc  | gagctgggccc | gacgaggggt  | cgagttcgcc  | 120 |
| gtcttccccg | agaccgtcat | tccctactat  | ccctacttct  | ctttcgtgca  | gccccctac   | 180 |
| accaggcca  | ccgaacacct | gcgcttgctc  | gaggaatcgg  | tgaccgtgcc  | ctccgccgaa  | 240 |
| accgacgcga | tcgccaaggc | cgctcgcgag  | gcggggcatgg | tcgtctccat  | cggcgtcaac  | 300 |
| gagcgcgacg | gcggaaccat | ctacaacacc  | caactcctct  | tcgacgccga  | cggcactctc  | 360 |
| atccagcgcc | gccgcaagat | caccccacc   | tatcacgaac  | gcatgggtctg | ggggcaggga  | 420 |
| gacggctcag | gtctgcgcg  | cgctcgacagt | gcggtcggcc  | gcatcggcca  | gctcgccctgc | 480 |
| tgggagcact | accagccact | ggcccgttac  | gccctcatcg  | ctgacggcga  | gcagatccac  | 540 |
| gccgcgatgt | accccggcgc | cttcggcggc  | gatctgttcg  | ccgagcagat  | cgaggtcaac  | 600 |



|            |            |            |             |            |            |     |
|------------|------------|------------|-------------|------------|------------|-----|
| atccgccagc | acgccctgga | atccgccagc | ttcgtcgtca  | acgccaccgc | ctggctcgac | 660 |
| gccgatcagc | aggcccagat | cgccaaggac | accggaggcc  | cgggtcccgc | cttctccggt | 720 |
| ggcttcttca | ccgccatcgt | cgaccccga  | ggccgtatca  | tcggcgaccc | cctcaccagc | 780 |
| ggcgaaggcg | aagtgatcgc | cgacctcgat | ctcgtctctca | tcaaccgccg | caagcgcctc | 840 |
| atggacgcca | gtggacacta | ccagccgccc | gaaattctta  | gcttcacatt | gaccggtgca | 900 |
| ccggcgcctt | atgtcaagag | cgcggcgtgc | cggggaaccc  | cgggtacgac | cgtggccgag | 960 |
| gagggacggt | ccgcttag   |            |             |            |            | 978 |

&lt;210&gt; 138

&lt;211&gt; 325

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 138

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Ile | Val | Lys | Ala | Ala | Ala | Val | Gln | Ile | Ser | Pro | Val | Leu | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Ser | Arg | Ala | Gly | Thr | Val | Asp | Lys | Val | Val | Ala | Lys | Ile | Arg | Glu | Leu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Gly | Arg | Arg | Gly | Val | Glu | Phe | Ala | Val | Phe | Pro | Glu | Thr | Val | Ile | Pro |
|     |     |     | 35  |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Tyr | Tyr | Pro | Tyr | Phe | Ser | Phe | Val | Gln | Pro | Pro | Tyr | Thr | Gln | Ala | Thr |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Glu | His | Leu | Arg | Leu | Leu | Glu | Glu | Ser | Val | Thr | Val | Pro | Ser | Ala | Glu |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Thr | Asp | Ala | Ile | Ala | Lys | Ala | Ala | Arg | Glu | Ala | Gly | Met | Val | Val | Ser |
|     |     |     |     | 85  |     |     |     | 90  |     |     |     |     |     | 95  |     |
| Ile | Gly | Val | Asn | Glu | Arg | Asp | Gly | Gly | Thr | Ile | Tyr | Asn | Thr | Gln | Leu |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |
| Leu | Phe | Asp | Ala | Asp | Gly | Thr | Leu | Ile | Gln | Arg | Arg | Arg | Lys | Ile | Thr |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Pro | Thr | Tyr | His | Glu | Arg | Met | Val | Trp | Gly | Gln | Gly | Asp | Gly | Ser | Gly |
|     |     | 130 |     |     |     | 135 |     |     |     |     |     | 140 |     |     |     |
| Leu | Arg | Ala | Val | Asp | Ser | Ala | Val | Gly | Arg | Ile | Gly | Gln | Leu | Ala | Cys |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Trp | Glu | His | Tyr | Gln | Pro | Leu | Ala | Arg | Tyr | Ala | Leu | Ile | Ala | Asp | Gly |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Glu | Gln | Ile | His | Ala | Ala | Met | Tyr | Pro | Gly | Ala | Phe | Gly | Gly | Asp | Leu |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Phe | Ala | Glu | Gln | Ile | Glu | Val | Asn | Ile | Arg | Gln | His | Ala | Leu | Glu | Ser |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Ala | Ser | Phe | Val | Val | Asn | Ala | Thr | Ala | Trp | Leu | Asp | Ala | Asp | Gln | Gln |
|     |     | 210 |     |     |     | 215 |     |     |     |     |     | 220 |     |     |     |
| Ala | Gln | Ile | Ala | Lys | Asp | Thr | Gly | Gly | Pro | Val | Pro | Ala | Phe | Ser | Gly |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Gly | Phe | Phe | Thr | Ala | Ile | Val | Asp | Pro | Glu | Gly | Arg | Ile | Ile | Gly | Asp |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Pro | Leu | Thr | Ser | Gly | Glu | Gly | Glu | Val | Ile | Ala | Asp | Leu | Asp | Leu | Ala |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     |     | 270 |     |
| Leu | Ile | Asn | Arg | Arg | Lys | Arg | Leu | Met | Asp | Ala | Ser | Gly | His | Tyr | Gln |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Pro | Pro | Glu | Ile | Leu | Ser | Phe | Thr | Leu | Thr | Gly | Ala | Pro | Ala | Pro | Tyr |
|     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Val | Lys | Ser | Ala | Ala | Cys | Arg | Gly | Thr | Pro | Gly | Thr | Thr | Val | Ala | Glu |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Glu | Gly | Arg | Ser | Ala |     |     |     |     |     |     |     |     |     |     |     |
|     |     |     |     | 325 |     |     |     |     |     |     |     |     |     |     |     |

&lt;210&gt; 139



WO 03/097810

PCT/US03/15712

<211> 999  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 139  
 atgaaaacaa cggttaccgt tgcctgcgtt caggccgccc ccgtatttat ggatttagaa 60  
 ggcaccgtag ataaaacaat caccctcatc tctgaagccg cacagaaagg cgcggagctc 120  
 atcgcttttc cggagacctg gatacccggt taccogtggg tcttatggct gaactcgccc 180  
 gccacaaata tgcccctggg ttatcagtat catcagaact ctctgggtgct ggacagtacc 240  
 caggcgaagc gaattgcgga tgcggcacgg cagaataaca tcaactgtcgc tctgggcttc 300  
 agcgaacgcg atcatggaag cctctatatc gcacagtggc tgattggcag cgacggggag 360  
 accattggca tccggcgcaa gctcaaggcc acgcacgtgg agcgtacgct gttcggcgaa 420  
 agcgacggct cctccctgac cacctgggag acacctctgg gtaacgtcgg ggccctctgc 480  
 tgctggggagc acctgcagcc gctgtcccgc tatgcaatgt attcccagca tgaggagatc 540  
 cacatcgctg cctggcccag tttcagtcctc tacaccagtg caacggccgc actgggtcct 600  
 gacgtcaata cggcggtctc acgcctctat gccgcggagg ggcagtgctt cgtgatagcc 660  
 ccgtgtgccg tggtttctga tgaaatgatt gatttactct gtcctgatga tgaccggaga 720  
 gcgttactca gtgccggagg gggacatgcc cgtatttacg gcccgacgg aagagaactc 780  
 gtcaccctc tcggggaaaa tgaggaagga ctgcttatcg ctgagctcga ctctgctgcg 840  
 attacctttg ccaaactggc ggcagacccg gttggccact attcccgtcc tgacgtgacc 900  
 cgcctccttt ttaatccttc agccaacaag actgtgatta aacgacattc gcctcctgag 960  
 ttaattgccg agcagactgc agaagaagag gaggagtag 999

<210> 140  
 <211> 332  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 140  
 Met Lys Thr Thr Val Thr Val Ala Cys Val Gln Ala Ala Pro Val Phe  
 1 5 10 15  
 Met Asp Leu Glu Gly Thr Val Asp Lys Thr Ile Thr Leu Ile Ser Glu  
 20 25 30  
 Ala Ala Gln Lys Gly Ala Glu Leu Ile Ala Phe Pro Glu Thr Trp Ile  
 35 40 45  
 Pro Gly Tyr Pro Trp Phe Leu Trp Leu Asn Ser Pro Ala Thr Asn Met  
 50 55 60  
 Pro Leu Val Tyr Gln Tyr His Gln Asn Ser Leu Val Leu Asp Ser Thr  
 65 70 75 80  
 Gln Ala Lys Arg Ile Ala Asp Ala Ala Arg Gln Asn Asn Ile Thr Val  
 85 90 95  
 Ala Leu Gly Phe Ser Glu Arg Asp His Gly Ser Leu Tyr Ile Ala Gln  
 100 105 110  
 Trp Leu Ile Gly Ser Asp Gly Glu Thr Ile Gly Ile Arg Arg Lys Leu  
 115 120 125  
 Lys Ala Thr His Val Glu Arg Thr Leu Phe Gly Glu Ser Asp Gly Ser  
 130 135 140  
 Ser Leu Thr Thr Trp Glu Thr Pro Leu Gly Asn Val Gly Ala Leu Cys  
 145 150 155 160  
 Cys Trp Glu His Leu Gln Pro Leu Ser Arg Tyr Ala Met Tyr Ser Gln  
 165 170 175  
 His Glu Glu Ile His Ile Ala Ala Trp Pro Ser Phe Ser Leu Tyr Thr  
 180 185 190  
 Ser Ala Thr Ala Ala Leu Gly Pro Asp Val Asn Thr Ala Ala Ser Arg  
 195 200 205



Leu Tyr Ala Ala Glu Gly Gln Cys Phe Val Ile Ala Pro Cys Ala Val  
 210 215 220  
 Val Ser Asp Glu Met Ile Asp Leu Leu Cys Pro Asp Asp Asp Arg Arg  
 225 230 235 240  
 Ala Leu Leu Ser Ala Gly Gly Gly His Ala Arg Ile Tyr Gly Pro Asp  
 245 250 255  
 Gly Arg Glu Leu Val Thr Pro Leu Gly Glu Asn Glu Glu Gly Leu Leu  
 260 265 270  
 Ile Ala Glu Leu Asp Ser Ala Ala Ile Thr Phe Ala Lys Leu Ala Ala  
 275 280 285  
 Asp Pro Val Gly His Tyr Ser Arg Pro Asp Val Thr Arg Leu Leu Phe  
 290 295 300  
 Asn Pro Ser Ala Asn Lys Thr Val Ile Lys Arg His Ser Pro Pro Glu  
 305 310 315 320  
 Leu Ile Ala Glu Gln Thr Ala Glu Glu Glu Glu Glu  
 325 330

<210> 141  
 <211> 1026  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 141  
 atggtgttca aggcagcgac tgttcatgca gctccggtat tcatggacaa ggaagcgtcg 60  
 atagataagg ctatcgacct catcaagaag gccggtcagg aagggattaa gcttctgggt 120  
 tttccgaaa cgtttattcc gggctatccg tattttatcg aatgctatcc gccgcttgcg 180  
 caggtggaag cgctcgcca gtacactgac gcttccgtgg agatcgacgg cccggaagtc 240  
 acccggttc agcaggtagc caaggcggca ggcgttgacg tcgtcatggg catcagcgaa 300  
 cgaatggctg agaccgaac ctgcttcaac tcgcaggtgt tcattgacgt cgacggcacg 360  
 ctgctcggcg tgcacgcaa gctgcagccg acttatgccg agcgcaaggt atgggcacag 420  
 ggcggtggtt atacgctgag gacctacaag agctcgcttg gcgtgctcgg cggctctgcc 480  
 tgctgggagc acacgatgaa cctcgcgcgg caggccctga tcatgcagag cgagcagatc 540  
 catgcggctg catggcccgg actatcgacg atgcgaggtt tcgagcccgt ggccgatatc 600  
 cagatcgacg ccatgatgaa gactcacgcg cttaccgcac agtgctgggt gctttcggcc 660  
 ggcaatcccg tcgaccggac ctgcctcgac tggatggaaa agaacatcgg accgcaggat 720  
 tacgtcaccg agggcggcgg atggagcgcc gttatccatc cggtcaacag ctatctcggc 780  
 ggccctcaca cgggccttga ggaaaagctg gtcgtcggcg agatcaatct ggacgatctc 840  
 aagttcgtca aagtctggct cgacagcaaa gggcactatg ctcggccgga aatcctgaaa 900  
 cttggcgtca accaaaagca gatttggcct gatgaacatt tgctggcgcg gcaggatgtg 960  
 accgagttgc tggaggcgga tatcatcgaa tacccttgc aactgttgca agaccgcgcg 1020  
 caatag 1026

<210> 142  
 <211> 341  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 142  
 Met Val Phe Lys Ala Ala Thr Val His Ala Ala Pro Val Phe Met Asp  
 1 5 10 15  
 Lys Glu Ala Ser Ile Asp Lys Ala Ile Asp Leu Ile Lys Lys Ala Gly  
 20 25 30  
 Gln Glu Gly Ile Lys Leu Leu Val Phe Pro Glu Thr Phe Ile Pro Gly  
 35 40 45  
 Tyr Pro Tyr Phe Ile Glu Cys Tyr Pro Pro Leu Ala Gln Val Glu Ala



|                         |                         |                         |
|-------------------------|-------------------------|-------------------------|
| 50                      | 55                      | 60                      |
| Leu Ala Gln Tyr Thr Asp | Ala Ser Val Glu Ile Asp | Gly Pro Glu Val         |
| 65                      | 70                      | 75                      |
| Thr Arg Leu Gln Gln Val | Ala Lys Ala Ala Gly     | Val Ala Val Val Met     |
|                         | 85                      | 90                      |
| Gly Ile Ser Glu Arg Met | Ala Glu Thr Arg Thr     | Cys Phe Asn Ser Gln     |
|                         | 100                     | 105                     |
| Val Phe Ile Asp Val Asp | Gly Thr Leu Leu Gly     | Val His Arg Lys Leu     |
|                         | 115                     | 120                     |
| Gln Pro Thr Tyr Ala Glu | Arg Lys Val Trp Ala     | Gln Gly Gly Gly Tyr     |
|                         | 130                     | 135                     |
| Thr Leu Arg Thr Tyr Lys | Ser Ser Leu Gly Val     | Leu Gly Gly Leu Ala     |
|                         | 145                     | 150                     |
| Cys Trp Glu His Thr Met | Asn Leu Ala Arg         | Gln Ala Leu Ile Met Gln |
|                         | 165                     | 170                     |
| Ser Glu Gln Ile His Ala | Ala Ala Trp Pro         | Gly Leu Ser Thr Met Arg |
|                         | 180                     | 185                     |
| Gly Phe Glu Pro Val Ala | Asp Ile Gln Ile Asp     | Ala Met Met Lys Thr     |
|                         | 195                     | 200                     |
| His Ala Leu Thr Ala Gln | Cys Trp Val Leu Ser     | Ala Gly Asn Pro Val     |
|                         | 210                     | 215                     |
| Asp Arg Thr Cys Leu Asp | Trp Met Glu Lys Asn     | Ile Gly Pro Gln Asp     |
|                         | 225                     | 230                     |
| Tyr Val Thr Glu Gly Gly | Gly Trp Ser Ala Val     | Ile His Pro Phe Asn     |
|                         | 245                     | 250                     |
| Ser Tyr Leu Gly Gly Pro | His Thr Gly Leu Glu     | Glu Lys Leu Val Val     |
|                         | 260                     | 265                     |
| Gly Glu Ile Asn Leu Asp | Asp Leu Lys Phe Val     | Lys Val Trp Leu Asp     |
|                         | 275                     | 280                     |
| Ser Lys Gly His Tyr Ala | Arg Pro Glu Ile Leu     | Lys Leu Gly Val Asn     |
|                         | 290                     | 295                     |
| Gln Lys Gln Ile Trp Pro | Asp Glu His Leu Leu     | Ala Arg Gln Asp Val     |
|                         | 305                     | 310                     |
| Thr Glu Leu Leu Glu Ala | Asp Ile Ile Glu Tyr     | Pro Leu Gln Leu Leu     |
|                         | 325                     | 330                     |
| Gln Asp Arg Ala Gln     |                         |                         |
|                         | 340                     |                         |

<210> 143  
 <211> 1122  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 143

|            |            |            |            |            |             |     |
|------------|------------|------------|------------|------------|-------------|-----|
| atgacgatca | ttgcaggcgc | ggttcacgac | gcgcccggat | tcatggatgt | cgatgccact  | 60  |
| atcgacaagg | catgcgaaat | cattcgcaag | gcaggcaaag | acggaatcga | gcttctcgtc  | 120 |
| ttccctgagg | ttttcgtacc | cggtacccc  | tacttcacgc | agtgcctacc | gaccttgaac  | 180 |
| caaaccgctg | cgctggccgc | ctatacggat | gcctcgatcg | aggttccagg | cccgggaagtc | 240 |
| cggcgcttgc | aggtggccgc | acatcaggcc | ggcgtgatgg | ttgtgatggg | cgtgagcgag  | 300 |
| cgtctgcgcg | gatctcgcac | ctgcttcaac | agccagggtg | tcatcgaccg | tgacggcacc  | 360 |
| ttgctggggc | tgcaccgcaa | actccagccg | acctatgtcg | agcgcacgtg | ctggggccag  | 420 |
| ggcggcggac | acaccctcaa | ggtattcgac | agcacactgg | gcaagggtgg | cggactggcc  | 480 |
| tgctggggag | acacgatgaa | cctcgcgcg  | catgcgttga | tcgcccagg  | tatccagatc  | 540 |
| catgccgccc | cctggcctgg | gctttcgaca | atggccgggt | tcgaagcgg  | ggctgacgtc  | 600 |
| cagatcgacg | cgatgatgaa | aactcatgcg | ttgagcgcg  | aatgctttgt | cgtatcggcc  | 660 |
| gcaaaccctg | tggatcagac | ctgcctggag | tggatggaga | aacacctcgg | cccgcagcaa  | 720 |
| ctcgttaccg | ccggcgagg  | ctggtcggca | atcgtccatc | ctttctgtgg | ttatatcgcc  | 780 |
| gcccctcaca | ccggtgccga | ggagaagggt | ctggtaggcg | aatcaatct  | ggacgacctc  | 840 |



```

aagcaggtca aggtatgggt tgattccgca ggtcattatg cgcgcccgga agtcgtgcaa 900
ttgcgcgacg ccctggagag ccgtggcaat tatcgcggtg cgctgacccg cgacgccgac 960
accttcgtgc cgctggaaga ccgcgtgcgc tttgcgcgcc agcagaacgc cgacctcttc 1020
atctcgatcc acgccgacgc caacgccaac cacgatgcgc gcggggctgg cttcacttcg 1080
aaggttgaaa acctttccac gggcatttta ccaggcgatt ga 1122

```

<210> 144  
 <211> 373  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 144

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Thr | Ile | Ile | Ala | Gly | Ala | Val | His | Ala | Ala | Pro | Val | Phe | Met | Asp | 1   | 5   | 10  | 15  |
| Val | Asp | Ala | Thr | Ile | Asp | Lys | Ala | Cys | Glu | Ile | Ile | Arg | Lys | Ala | Gly | 20  | 25  | 30  |     |
| Lys | Asp | Gly | Ile | Glu | Leu | Leu | Val | Phe | Pro | Glu | Val | Phe | Val | Pro | Gly | 35  | 40  | 45  |     |
| Tyr | Pro | Tyr | Phe | Ile | Glu | Cys | Tyr | Pro | Thr | Leu | Asn | Gln | Thr | Ala | Ala | 50  | 55  | 60  |     |
| Leu | Ala | Ala | Tyr | Thr | Asp | Ala | Ser | Ile | Glu | Val | Pro | Gly | Pro | Glu | Val | 65  | 70  | 75  | 80  |
| Arg | Arg | Leu | Gln | Val | Ala | Ala | His | Gln | Ala | Gly | Val | Met | Val | Val | Met | 85  | 90  | 95  |     |
| Gly | Val | Ser | Glu | Arg | Leu | Arg | Gly | Ser | Arg | Thr | Cys | Phe | Asn | Ser | Gln | 100 | 105 | 110 |     |
| Val | Phe | Ile | Asp | Arg | Asp | Gly | Thr | Leu | Leu | Gly | Val | His | Arg | Lys | Leu | 115 | 120 | 125 |     |
| Gln | Pro | Thr | Tyr | Val | Glu | Arg | Ile | Val | Trp | Gly | Gln | Gly | Gly | Gly | His | 130 | 135 | 140 |     |
| Thr | Leu | Lys | Val | Phe | Asp | Ser | Thr | Leu | Gly | Lys | Val | Gly | Gly | Leu | Ala | 145 | 150 | 155 | 160 |
| Cys | Trp | Glu | His | Thr | Met | Asn | Leu | Ala | Arg | His | Ala | Leu | Ile | Ala | Gln | 165 | 170 | 175 |     |
| Gly | Ile | Gln | Ile | His | Ala | Ala | Ala | Trp | Pro | Gly | Leu | Ser | Thr | Met | Ala | 180 | 185 | 190 |     |
| Gly | Phe | Glu | Ala | Val | Ala | Asp | Val | Gln | Ile | Asp | Ala | Met | Met | Lys | Thr | 195 | 200 | 205 |     |
| His | Ala | Leu | Ser | Ala | Gln | Cys | Phe | Val | Val | Ser | Ala | Ala | Asn | Pro | Val | 210 | 215 | 220 |     |
| Asp | Gln | Thr | Cys | Leu | Glu | Trp | Met | Glu | Lys | His | Leu | Gly | Pro | Gln | Gln | 225 | 230 | 235 | 240 |
| Leu | Val | Thr | Ala | Gly | Gly | Gly | Trp | Ser | Ala | Ile | Val | His | Pro | Phe | Cys | 245 | 250 | 255 |     |
| Gly | Tyr | Ile | Ala | Ala | Pro | His | Thr | Gly | Ala | Glu | Glu | Lys | Val | Leu | Val | 260 | 265 | 270 |     |
| Gly | Glu | Ile | Asn | Leu | Asp | Asp | Leu | Lys | Gln | Val | Lys | Val | Trp | Val | Asp | 275 | 280 | 285 |     |
| Ser | Ala | Gly | His | Tyr | Ala | Arg | Pro | Glu | Val | Val | Gln | Leu | Arg | Asp | Ala | 290 | 295 | 300 |     |
| Leu | Glu | Ser | Arg | Gly | Asn | Tyr | Arg | Val | Ala | Leu | Thr | Arg | Asp | Ala | Asp | 305 | 310 | 315 | 320 |
| Thr | Phe | Val | Pro | Leu | Glu | Asp | Arg | Val | Arg | Phe | Ala | Arg | Gln | Gln | Asn | 325 | 330 | 335 |     |
| Ala | Asp | Leu | Phe | Ile | Ser | Ile | His | Ala | Asp | Ala | Asn | Ala | Asn | His | Asp | 340 | 345 | 350 |     |
| Ala | Arg | Gly | Ala | Gly | Phe | Thr | Ser | Lys | Val | Glu | Asn | Leu | Ser | Thr | Gly | 355 | 360 | 365 |     |



WO 03/097810

PCT/US03/15712

Ile Leu Pro Gly Asp  
370

<210> 145  
<211> 1014  
<212> DNA  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 145  
atgggcatca cccacccgaa ctacaaggct gcagtggctc aggctgcgcc ggtctggttg 60  
aacctcgagg caacggtcga gaagacaatc aggtatatg aagaggcggc caaggctgga 120  
gcgaagctga tagcgtttcc ggaaacctgg attccgggct atccatggca catttggtac 180  
ggaacgcccg catgggcaat cggtaagggc ttcgtccagc gctatttcga caactcgctc 240  
agctatgaca gcccgctcgc gcggcagatc gctgacgccg cagcaaagag caagatcacg 300  
gttggttctc gcctctccga gcgcgacggg ggaagcctat acatcgcgca atggctgatc 360  
ggaccagatg gcgagaccat cgcgaagcgg cgcaagctgc gtccgaccca cgtcgagcgc 420  
acggtgttct gtgacgggtg cggcagccac atcgccgtgc atgaccgatc cgatctgggc 480  
cggctcgggg cgttgtgctg ctgggagcac gtgcagccgt tgacgaaatt cgcgatgtac 540  
gcgcagaacg agcaggttca cgtggcagca tggccgagct tctcgatgta cgaacccttt 600  
gcgcatgcgc tgggttggga gacgaacaac gcggtcagca aggtctacgc ggtcgaggga 660  
tcgtgcttct tgctcgctcc ctgtgccgtt atttcgcaag cgatgggtga cgagatgtgc 720  
gacactcccc acaagcgcgga gcttggtcac gccggcggcg gccacgcggg gatattacggc 780  
cctgacggaa gcccgctcgc agaaaagctc ggggaaaacg aagaggggct tctctacgcg 840  
acggtcaatc ttgctgcgat cgggggttgcc aagaatgccg cggatccggc cgggcactat 900  
tcgcgtccgg acgttctaag gctgctattc aacaagagcc cggcccgaag agtggagcat 960  
tttgcgctgc cgcacgagca gctcgagatc ggggcaggcc cgtctggcga ctga 1014

<210> 146  
<211> 337  
<212> PRT  
<213> Unknown

<220>  
<223> Obtained from an environmental sample

<400> 146  
Met Gly Ile Thr His Pro Asn Tyr Lys Val Ala Val Val Gln Ala Ala  
1 5 10 15  
Pro Val Trp Leu Asn Leu Glu Ala Thr Val Glu Lys Thr Ile Arg Tyr  
20 25 30  
Ile Glu Glu Ala Ala Lys Ala Gly Ala Lys Leu Ile Ala Phe Pro Glu  
35 40 45  
Thr Trp Ile Pro Gly Tyr Pro Trp His Ile Trp Ile Gly Thr Pro Ala  
50 55 60  
Trp Ala Ile Gly Lys Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu  
65 70 75 80  
Ser Tyr Asp Ser Pro Leu Ala Arg Gln Ile Ala Asp Ala Ala Ala Lys  
85 90 95  
Ser Lys Ile Thr Val Val Leu Gly Leu Ser Glu Arg Asp Gly Gly Ser  
100 105 110  
Leu Tyr Ile Ala Gln Trp Leu Ile Gly Pro Asp Gly Glu Thr Ile Ala  
115 120 125  
Lys Arg Arg Lys Leu Arg Pro Thr His Val Glu Arg Thr Val Phe Gly  
130 135 140  
Asp Gly Asp Gly Ser His Ile Ala Val His Asp Arg Ser Asp Leu Gly  
145 150 155 160  
Arg Leu Gly Ala Leu Cys Cys Trp Glu His Val Gln Pro Leu Thr Lys  
165 170 175



Phe Ala Met Tyr Ala Gln Asn Glu Gln Val His Val Ala Ala Trp Pro  
 180 185 190  
 Ser Phe Ser Met Tyr Glu Pro Phe Ala His Ala Leu Gly Trp Glu Thr  
 195 200 205  
 Asn Asn Ala Val Ser Lys Val Tyr Ala Val Glu Gly Ser Cys Phe Val  
 210 215 220  
 Leu Ala Pro Cys Ala Val Ile Ser Gln Ala Met Val Asp Glu Met Cys  
 225 230 235 240  
 Asp Thr Pro Asp Lys Arg Glu Leu Val His Ala Gly Gly Gly His Ala  
 245 250 255  
 Val Ile Tyr Gly Pro Asp Gly Ser Pro Leu Ala Glu Lys Leu Gly Glu  
 260 265 270  
 Asn Glu Glu Gly Leu Leu Tyr Ala Thr Val Asn Leu Ala Ala Ile Gly  
 275 280 285  
 Val Ala Lys Asn Ala Ala Asp Pro Ala Gly His Tyr Ser Arg Pro Asp  
 290 295 300  
 Val Leu Arg Leu Leu Phe Asn Lys Ser Pro Ala Arg Arg Val Glu His  
 305 310 315 320  
 Phe Ala Leu Pro His Glu Gln Leu Glu Ile Gly Ala Gly Pro Ser Gly  
 325 330 335  
 Asp

<210> 147  
 <211> 1098  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 147  
 atgaccacgc acgagaccac tgcccggagg ctggcagctg tgcattgccgc gcctgtgttc 60  
 atggacaccg acgcgaccat cgacaagggtg atcggcttcg tcgaacaggc cggccgcgaa 120  
 ggcatcgaac tcttggtgtt ccccagagacc ttcgtgctg gttacccta ctggatcgag 180  
 tgctatccgc cgctgcagca ggtggccgcc aacgcgcagt acacggacgc ctccgtcgag 240  
 gtgcctggtc cggagatcaa gcgggtgcag gcggcctgtg cccgcgctgg cgtcgaagtc 300  
 gtcctcggcg tcagcgagcg actcaggggt accaggacat gcttcaactc ccaggtgttc 360  
 atcgacgccg acgggagcct gctcggcgtg caccgcaagc tgcagccgac gtacgtggag 420  
 cgcattcgtg gggcccaggg cggaggcgcg accctgtcgg tgttcggctc ccgctccggc 480  
 cggatcggcg gtctggcctg ctgggagcac acgatgaacc tggctcgtca ggcactgctt 540  
 gagcaggagc agcagatcca cgcggcggcg tggcctgccc tgtcgacgat ggcgggggtt 600  
 gagaccgtcg cggacgcccc gatcgaggcc atgatgaaga cccatgcgct cacggcacag 660  
 gtgttcgtca tctgcgcgtc caaccgggtc gacggcactt gcctggaatg gatgcgggac 720  
 aacctcgggtg aacagaagtt cgtgaccgcc ggaggggggt ggtccgcggt catccacccc 780  
 ttcaactcct tctcggcggg gccgcatacc ggtttggagg agaagctcgt cagcgcgacg 840  
 atcgacttct ccgacatccg cttgggtcaag gcctgggttg attcgaaggg gcactacgcg 900  
 cggcccaggg tcttgcgact cgcggctgcac cgcaagccac tgtggcacga cgagtgcgag 960  
 gtgcccgggac aggcgcaggt acgcacccgc gctgcttctc tggcagtgca ggagcaccgc 1020  
 gtggtgctgc ctcagggggc ggcgcggccc gctccgcaag actgggacac ctctgcggcg 1080  
 caggagctga cttcctga 1098

<210> 148  
 <211> 365  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 148



Met Thr Gln His Glu Thr Thr Ala Arg Arg Leu Ala Ala Val His Ala  
 1 5 10 15  
 Ala Pro Val Phe Met Asp Thr Asp Ala Thr Ile Asp Lys Val Ile Gly  
 20 25 30  
 Phe Val Glu Gln Ala Gly Arg Glu Gly Ile Glu Leu Leu Val Phe Pro  
 35 40 45  
 Glu Thr Phe Val Pro Gly Tyr Pro Tyr Trp Ile Glu Cys Tyr Pro Pro  
 50 55 60  
 Leu Gln Gln Val Ala Ala Asn Ala Gln Tyr Thr Asp Ala Ser Val Glu  
 65 70 75 80  
 Val Pro Gly Pro Glu Ile Lys Arg Val Gln Ala Ala Cys Ala Arg Ala  
 85 90 95  
 Gly Val Glu Val Val Leu Gly Val Ser Glu Arg Leu Arg Gly Thr Arg  
 100 105 110  
 Thr Cys Phe Asn Ser Gln Val Phe Ile Asp Ala Asp Gly Ser Leu Leu  
 115 120 125  
 Gly Val His Arg Lys Leu Gln Pro Thr Tyr Val Glu Arg Ile Val Trp  
 130 135 140  
 Ala Gln Gly Gly Gly Ala Thr Leu Ser Val Phe Gly Ser Arg Ser Gly  
 145 150 155 160  
 Arg Ile Gly Gly Leu Ala Cys Trp Glu His Thr Met Asn Leu Ala Arg  
 165 170 175  
 Gln Ala Leu Leu Glu Gln Glu Gln Ile His Ala Ala Ala Trp Pro  
 180 185 190  
 Ala Leu Ser Thr Met Ala Gly Phe Glu Thr Val Ala Asp Ala Gln Ile  
 195 200 205  
 Glu Ala Met Met Lys Thr His Ala Leu Thr Ala Gln Val Phe Val Ile  
 210 215 220  
 Cys Ala Ser Asn Pro Val Asp Gly Thr Cys Leu Glu Trp Met Arg Asp  
 225 230 235 240  
 Asn Leu Gly Glu Gln Lys Phe Val Thr Ala Gly Gly Gly Trp Ser Ala  
 245 250 255  
 Val Ile His Pro Phe Asn Ser Phe Leu Gly Gly Pro His Thr Gly Leu  
 260 265 270  
 Glu Glu Lys Leu Val Ser Ala Thr Ile Asp Phe Ser Asp Ile Arg Leu  
 275 280 285  
 Val Lys Ala Trp Val Asp Ser Lys Gly His Tyr Ala Arg Pro Glu Val  
 290 295 300  
 Leu Arg Leu Ala Val Asp Arg Lys Pro Leu Trp His Asp Glu Cys Glu  
 305 310 315 320  
 Val Pro Gly Gln Ala Gln Val Arg Thr Arg Ala Ala Ser Leu Ala Val  
 325 330 335  
 Gln Glu His Pro Val Val Leu Pro Gln Gly Ala Ala Arg Pro Ala Pro  
 340 345 350  
 Gln Asp Trp Asp Thr Ser Ala Ala Gln Glu Leu Thr Ser  
 355 360 365

&lt;210&gt; 149

&lt;211&gt; 942

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 149

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| atgacgaagc | ttgagaaggt | ggtcgcggcg | gcggtccagg | cgacgccgga | gttcctcgac | 60  |
| cgcgaggcga | ccgtcgagaa | ggccgtgctg | ctgatcaagg | aagcggccgg | ggagggcgcc | 120 |
| ggcctgatcg | tggtccccga | gacgttcac  | ccgacgtacc | cggactgggt | ctggcgcgcg | 180 |
| ccggcctggg | acggcccatc | cgcgacctg  | tacgcaatgc | tgctggagaa | cgcggtggag | 240 |
| atccccgggc | cggtgacgga | gaccctgggg | aaggcggcga | agcaggccaa | ggccttcgtg | 300 |



```

tcgatgggcg tcaacgagcg cgagccgggc ggcgggacga tctacaacac gcaggtcacg 360
ttcggacccg acgggagcgt gctcggcaag caccgcaagc tgatgccgac cggcggcgag 420
cgcttggtgt gggggatggg cgacgggtcg atgctccagg tctatgacac gccgttcggc 480
cgctggggcg ggctgatctg ctgggagaaac tacatgccgc tcgcgcgcta ctcgatgtac 540
gccaagggcg tggacgtcta cgttgccgccc acgtgggaca acagcgacat gtgggtggcg 600
acgtcccgcc acatcgccaa ggagggggcg ctgtacgtga tcggcgtggc gccgctgctg 660
cgcggttcgg acgtccccga cgacgtgccg gggaaggccg agctgtgggg cggcgatgac 720
gactggatgt cgcgcggctt ctccaccatc gtcgcgccgg gcggcgaggt gctggccggt 780
ccgtgacgg aggaggaagg catcctctac gcggagatcg acccggcgag agcccgttcg 840
tcacggcacc agttcgatcc ggtggggcac tactcgccgc ccgacgtgtt tcggctcgtc 900
gtggacgagt cgcccaagcc ccagacgtcc ggcccgggct ag 942

```

&lt;210&gt; 150

&lt;211&gt; 313

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 150

```

Met Thr Lys Leu Glu Lys Val Val Ala Ala Ala Val Gln Ala Thr Pro
 1          5          10          15
Glu Phe Leu Asp Arg Glu Ala Thr Val Glu Lys Ala Val Arg Leu Ile
 20          25          30
Lys Glu Ala Ala Gly Glu Gly Ala Gly Leu Ile Val Phe Pro Glu Thr
 35          40          45
Phe Ile Pro Thr Tyr Pro Asp Trp Val Trp Arg Ala Pro Ala Trp Asp
 50          55          60
Gly Pro Ser Ala Asp Leu Tyr Ala Met Leu Leu Glu Asn Ala Val Glu
 65          70          75          80
Ile Pro Gly Pro Val Thr Glu Thr Leu Gly Lys Ala Ala Lys Gln Ala
 85          90          95
Lys Ala Phe Val Ser Met Gly Val Asn Glu Arg Glu Pro Gly Gly Gly
 100          105          110
Thr Ile Tyr Asn Thr Gln Val Thr Phe Gly Pro Asp Gly Ser Val Leu
 115          120          125
Gly Lys His Arg Lys Leu Met Pro Thr Gly Gly Glu Arg Leu Val Trp
 130          135          140
Gly Met Gly Asp Gly Ser Met Leu Gln Val Tyr Asp Thr Pro Phe Gly
 145          150          155          160
Arg Leu Gly Gly Leu Ile Cys Trp Glu Asn Tyr Met Pro Leu Ala Arg
 165          170          175
Tyr Ser Met Tyr Ala Lys Gly Val Asp Val Tyr Val Ala Pro Thr Trp
 180          185          190
Asp Asn Ser Asp Met Trp Val Ala Thr Leu Arg His Ile Ala Lys Glu
 195          200          205
Gly Arg Leu Tyr Val Ile Gly Val Ala Pro Leu Leu Arg Gly Ser Asp
 210          215          220
Val Pro Asp Asp Val Pro Gly Lys Ala Glu Leu Trp Gly Gly Asp Asp
 225          230          235          240
Asp Trp Met Ser Arg Gly Phe Ser Thr Ile Val Ala Pro Gly Gly Glu
 245          250          255
Val Leu Ala Gly Pro Leu Thr Glu Glu Glu Gly Ile Leu Tyr Ala Glu
 260          265          270
Ile Asp Pro Ala Arg Ala Arg Ser Ser Arg His Gln Phe Asp Pro Val
 275          280          285
Gly His Tyr Ser Arg Pro Asp Val Phe Arg Leu Val Val Asp Glu Ser
 290          295          300
Pro Lys Pro Gln Thr Ser Gly Pro Gly
 305          310

```



<210> 151  
 <211> 993  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 151  
 atgagagtcg ttaaagccgc cgcggtccaa ctgaaaccag tcctttatag ccgtgagggg 60  
 acagtcgata acgtcgtcaa gaagatccac gagctgggcc aacaaggagt gcagttcgca 120  
 acgttcccgg aaaccgtggt gccttactat ccgtactttt cgatcgtgca gtccggctat 180  
 caaatccttg ccggcgggtga gttcctaaag ctgcttgatc agtcagtac cgtgccatct 240  
 cttgccaccg aagcgatcgg cgaggcctgc aggcaagcgg gcgtcgttgt ctccatcggc 300  
 gtcaacgagc gtgacggggg aactctgtac aatacgcaac ttctctttga tgccgacggc 360  
 acgttgattc aaagacgacg caagatcacg cccacccatt acgagcgcac ggtctggggc 420  
 cagggcgatg gctcaggttt acgggcgggt gacagcaagg tcgcgcgcac tgggtcaactg 480  
 gcttggtttg agcactacaa cccgcttgcg cgttacgcca tgatggccga tggcgagcaa 540  
 atccactctg cgatgttccc gggctccatg ttccggcgat cgttttcaga gaaggtggaa 600  
 atcaacgtaa ggcagcatgc aatggagtct ggatgctttg tcgtctgcgc tacggcctgg 660  
 ctggatgccg accaacaggc acaaatcatg aaggacacag gctgcgagat cgggtccgatc 720  
 tcgggcgggt gcttcaccgc tatcgtgaca cccgacggga cgctgatagg cgaacccatc 780  
 cactcgggcg aaggcgtttg tattgccgac ctcgatttca agctcatcga caagcggaag 840  
 cacgtgggtg acacgcgcgg ccactacagc cggccagaat tgctcagcct cctaattgat 900  
 cggactccca cggcacacat acacgaacgg aacgagcaac cgaagtcggc cgttgagcaa 960  
 gactcgcaga atgtattcac cgctattgct taa 993

<210> 152  
 <211> 330  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 152  
 Met Arg Val Val Lys Ala Ala Ala Val Gln Leu Lys Pro Val Leu Tyr  
 1 5 10 15  
 Ser Arg Glu Gly Thr Val Asp Asn Val Val Lys Lys Ile His Glu Leu  
 20 25 30  
 Gly Gln Gln Gly Val Gln Phe Ala Thr Phe Pro Glu Thr Val Val Pro  
 35 40 45  
 Tyr Tyr Pro Tyr Phe Ser Ile Val Gln Ser Gly Tyr Gln Ile Leu Ala  
 50 55 60  
 Gly Gly Glu Phe Leu Lys Leu Leu Asp Gln Ser Val Thr Val Pro Ser  
 65 70 75 80  
 Leu Ala Thr Glu Ala Ile Gly Glu Ala Cys Arg Gln Ala Gly Val Val  
 85 90 95  
 Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Leu Tyr Asn Thr  
 100 105 110  
 Gln Leu Leu Phe Asp Ala Asp Gly Thr Leu Ile Gln Arg Arg Arg Lys  
 115 120 125  
 Ile Thr Pro Thr His Tyr Glu Arg Met Val Trp Gly Gln Gly Asp Gly  
 130 135 140  
 Ser Gly Leu Arg Ala Val Asp Ser Lys Val Ala Arg Ile Gly Gln Leu  
 145 150 155 160  
 Ala Cys Phe Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Met Met Ala  
 165 170 175  
 Asp Gly Glu Gln Ile His Ser Ala Met Phe Pro Gly Ser Met Phe Gly  
 180 185 190



Asp Ala Phe Ser Glu Lys Val Glu Ile Asn Val Arg Gln His Ala Met  
 195 200 205  
 Glu Ser Gly Cys Phe Val Val Cys Ala Thr Ala Trp Leu Asp Ala Asp  
 210 215 220  
 Gln Gln Ala Gln Ile Met Lys Asp Thr Gly Cys Glu Ile Gly Pro Ile  
 225 230 235 240  
 Ser Gly Gly Cys Phe Thr Ala Ile Val Thr Pro Asp Gly Thr Leu Ile  
 245 250 255  
 Gly Glu Pro Ile His Ser Gly Glu Gly Val Cys Ile Ala Asp Leu Asp  
 260 265 270  
 Phe Lys Leu Ile Asp Lys Arg Lys His Val Val Asp Thr Arg Gly His  
 275 280 285  
 Tyr Ser Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Thr Pro Thr  
 290 295 300  
 Ala His Ile His Glu Arg Asn Glu Gln Pro Lys Ser Ala Val Glu Gln  
 305 310 315 320  
 Asp Ser Gln Asn Val Phe Thr Ala Ile Ala  
 325 330

<210> 153  
 <211> 1074  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 153  
 atgccaaacg caagaaagat tggttgagcc gtggcccaag ttgcacagga attcttcgac 60  
 actgaagcga atctcggtaa agcgatagcg gcgattcaca atgctgcgaa gcaaggcgca 120  
 gatatcgctg tcttcgccga atgctatattg ggccaatata catattgggc gcaattttac 180  
 gacaactctg ccaagaacta ttccaagggt tggacggccc tgtacgacgg tgcgatcact 240  
 gtgggtggcg atgaatgccg ggctattgct gctgcggcta gacagtccaa gattcatgtc 300  
 gtcattgggt gcaatgagct atccgaccga gccggcggcg caacgttata caacagcctc 360  
 ttgtttttcg accgaaaggg cgagttgatc ggtcgacacc ggaaattgat gccgtcgatg 420  
 cacgagcggt tgatccatgg cacaggcgac ggaagagact tgaatgttta cgataccgat 480  
 atcgggtatgt tgggtgggtt gatttgctgg gagcaccata tgctgctctc gaagtatgcc 540  
 atggcgacta tgggtgaaga agttcatgtt gcaagctggc ctgggatgtg gcgcggagga 600  
 gacgcggcaa tcggtgagag gatggtcgaa gcgatcttg gggcgccgtt tgtttgtgac 660  
 gccgaatttg cgatccgaga atatgcggca gagacaggaa atttcgttct aagcgcgtct 720  
 ggatattttc cgaaggacaa tatatccgat gagtggcgag aagcgattcc aaaccttcaa 780  
 gcgcagtggg ctgtgggagg gagttctatc gtggcaccgg ggggctccta tctggtccca 840  
 ccactcatta atgaggagaa gatcctctgc gccgaactcg atttcaatct caggcgtctt 900  
 tggaaagcct ggatcgatcc gattggtcac tattcgcgtc ccgatgttta tagcctgcaa 960  
 ctgcataacg ttgctgggag tgagtattcc tatcaggccg tagatttgaa gcgcacgcca 1020  
 aagccccaat cgctgtgggt agatgcgtcc gaggaagacg gtgcgctgaa ttga 1074

<210> 154  
 <211> 357  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 154  
 Met Pro Asn Ala Arg Lys Ile Val Gly Ala Val Ala Gln Val Ala Gln  
 1 5 10 15  
 Glu Phe Phe Asp Thr Glu Ala Asn Leu Gly Lys Ala Ile Ala Ala Ile  
 20 25 30  
 His Asn Ala Ala Lys Gln Gly Ala Asp Ile Val Val Phe Ala Glu Cys



|            |             |            |            |            |            |     |
|------------|-------------|------------|------------|------------|------------|-----|
| <400> 155  |             |            |            |            |            |     |
| atgggcatcg | aacatccgaa  | atacaaggtg | gccgtggtgc | aggccgcgcc | ggcctggctc | 60  |
| gatctcgacg | gctcgatcaa  | gaaggcgatt | gcgctgatcg | aggaagcggc | cgccaagggc | 120 |
| gctaagctga | tcgctttccc  | cgaaaccttc | attcccggct | atccctggca | catctggctg | 180 |
| gactcgccgg | cctgggcgat  | cgccgcgcgc | tttgtgcagc | gctacttcga | taactcgctg | 240 |
| gcctacgaca | gcccgcgaagc | cgaaaagctg | cgcgccgcgc | tcaagaaggc | caagctcact | 300 |
| gccgtgattg | gcctgtcggg  | gcgcgacggc | ggcagcctct | atatagcgca | atggctgatt | 360 |
| ggccctgatg | gcgagaccat  | cgcaaaacgc | agaaagctgc | ggccaacgca | cgcggaacgc | 420 |
| accgtttttg | gcgaggggtg  | cggcagcgac | cttgccgtgc | acgaccggcc | cggaatcggg | 480 |
| cggctgggag | cgctgtgctg  | ctgggagcac | ctgcaaccgc | tttcgaaata | cgcgatgtat | 540 |
| gcgcagaacg | aacaggtcca  | tgtcgcgtca | tggccgaqct | tctcgtctga | cgaccccttc | 600 |



```

gcgcgcggcgc tcggcgccga ggtcaacaat gcggcttccc gcgtctacgc ggtcgagggc 660
tcgtgcttcg tgctggcgcc gtgcgccacg gtttcgcaag ccatgatcga cgagctgtgt 720
gaccggccgg acaagcatgc gctgttgac gccggtggcg gacacgccgc gatttacggc 780
ccggacggca gctcgatcgc ggagaagctg ccgcaggacg cggagggcct gttgatcgcc 840
gagatcgatc tcggggcgat cgggggttgcc aagaatgcag ccgaccggc cggtcattat 900
tcgcggccgg acgtgacgcg actcctgctg aacaagaacc ggatgcgaag ggtcgaggag 960
tttgcgctgc cggtcgatcc ggtcgcaacg accgaggagg agcaagtcgc gacgccgtcg 1020
aggcccagcc aggccgcgta a 1041

```

&lt;210&gt; 156

&lt;211&gt; 346

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 156

```

Met Gly Ile Glu His Pro Lys Tyr Lys Val Ala Val Val Gln Ala Ala
 1          5          10          15
Pro Ala Trp Leu Asp Leu Asp Gly Ser Ile Lys Lys Ala Ile Ala Leu
 20          25          30
Ile Glu Glu Ala Ala Ala Lys Gly Ala Lys Leu Ile Ala Phe Pro Glu
 35          40          45
Thr Phe Ile Pro Gly Tyr Pro Trp His Ile Trp Leu Asp Ser Pro Ala
 50          55          60
Trp Ala Ile Gly Arg Gly Phe Val Gln Arg Tyr Phe Asp Asn Ser Leu
 65          70          75          80
Ala Tyr Asp Ser Pro Gln Ala Glu Lys Leu Arg Ala Ala Val Lys Lys
 85          90          95
Ala Lys Leu Thr Ala Val Ile Gly Leu Ser Glu Arg Asp Gly Gly Ser
 100         105         110
Leu Tyr Ile Ala Gln Trp Leu Ile Gly Pro Asp Gly Glu Thr Ile Ala
 115         120         125
Lys Arg Arg Lys Leu Arg Pro Thr His Ala Glu Arg Thr Val Phe Gly
 130         135         140
Glu Gly Asp Gly Ser Asp Leu Ala Val His Asp Arg Pro Gly Ile Gly
 145         150         155         160
Arg Leu Gly Ala Leu Cys Cys Trp Glu His Leu Gln Pro Leu Ser Lys
 165         170         175
Tyr Ala Met Tyr Ala Gln Asn Glu Gln Val His Val Ala Ser Trp Pro
 180         185         190
Ser Phe Ser Leu Tyr Asp Pro Phe Ala Pro Ala Leu Gly Ala Glu Val
 195         200         205
Asn Asn Ala Ala Ser Arg Val Tyr Ala Val Glu Gly Ser Cys Phe Val
 210         215         220
Leu Ala Pro Cys Ala Thr Val Ser Gln Ala Met Ile Asp Glu Leu Cys
 225         230         235         240
Asp Arg Pro Asp Lys His Ala Leu Leu His Ala Gly Gly Gly His Ala
 245         250         255
Ala Ile Tyr Gly Pro Asp Gly Ser Ser Ile Ala Glu Lys Leu Pro Gln
 260         265         270
Asp Ala Glu Gly Leu Leu Ile Ala Glu Ile Asp Leu Gly Ala Ile Gly
 275         280         285
Val Ala Lys Asn Ala Ala Asp Pro Ala Gly His Tyr Ser Arg Pro Asp
 290         295         300
Val Thr Arg Leu Leu Leu Asn Lys Asn Arg Met Arg Arg Val Glu Glu
 305         310         315         320
Phe Ala Leu Pro Val Asp Pro Val Ala Thr Glu Glu Glu Gln Val
 325         330         335
Ala Thr Pro Ser Arg Pro Ser Gln Ala Ala

```



340

345

<210> 157  
 <211> 1011  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 157  
 atgagagtcg ttaaagctgc tgcgggtccaa ctgaggtcccg tgctgtatag ccgtgagggga 60  
 acagtagaaa aggtcgttcg gaagatccac gagcttggcg atcaaggagt cgagttcgcc 120  
 acgttcccgg agaccgtagt gccctactat ccgtacttct cggccgtcca gacgcgatt 180  
 cagaacatgc acggcccggg gcacctgaag ttgctcgagc aatcggtgac cgtcccgtcg 240  
 cccgccaccg acgcgatcgg cgacgcctgc cgccacgccg gcgtcgtcgt ctcgatcggc 300  
 gtcaacgaac gcgatggcgg cacgatctac aacacgcagc tcctgttcga cgccgacggc 360  
 accttgatcc agcgccggcg aaagatcacg ccgaccttct acgaacgaat ggtctgggga 420  
 cagggtgacg gttcggggct gcgcgccgtc gacagccgcg taggacgcat cggccagctc 480  
 gcctgtttcg agcactacaa cccgctggcg cgctacgcca tgatggccga cggcgagcag 540  
 attcactccg cgatgtaccc cggctccatc tttggagacg cattcgcgca gaaaatcgag 600  
 atcaacatcc gccagcacgc gctcgagtcc ggtgcgttcg tcgtcaacgc caccgcctgg 660  
 ctcgatgccg accagcaggc gcggatcatg aaggataccg gctgcaccat cgaaccgatc 720  
 tcgggcgggtt gcttcaccgc catcgtcacc ccggacggga ccctgctggg cgaagcgata 780  
 cgttcggggg agggagtggg ggtcgccgat ctcgacttca cgctgatcga caggcgcaag 840  
 caagtgatgg actctcgtgg tcactacagt cggccggagt tgctcagcct tctgatcgac 900  
 cgcacaccca ccgcacacct acacgaacgc gaagcgcacc ccagagcaag tgaggactgg 960  
 caaggttccg agagtctgcg cgccatgcag gcctcggcac cgaaggtctg a 1011

<210> 158  
 <211> 336  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 158  
 Met Arg Val Val Lys Ala Ala Ala Val Gln Leu Ser Pro Val Leu Tyr  
 1 5 10 15  
 Ser Arg Glu Gly Thr Val Glu Lys Val Val Arg Lys Ile His Glu Leu  
 20 25 30  
 Gly Asp Gln Gly Val Glu Phe Ala Thr Phe Pro Glu Thr Val Val Pro  
 35 40 45  
 Tyr Tyr Pro Tyr Phe Ser Ala Val Gln Thr Pro Ile Gln Asn Met His  
 50 55 60  
 Gly Pro Glu His Leu Lys Leu Leu Glu Gln Ser Val Thr Val Pro Ser  
 65 70 75 80  
 Pro Ala Thr Asp Ala Ile Gly Asp Ala Cys Arg His Ala Gly Val Val  
 85 90 95  
 Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Ile Tyr Asn Thr  
 100 105 110  
 Gln Leu Leu Phe Asp Ala Asp Gly Thr Leu Ile Gln Arg Arg Arg Lys  
 115 120 125  
 Ile Thr Pro Thr Phe Tyr Glu Arg Met Val Trp Gly Gln Gly Asp Gly  
 130 135 140  
 Ser Gly Leu Arg Ala Val Asp Ser Arg Val Gly Arg Ile Gly Gln Leu  
 145 150 155 160  
 Ala Cys Phe Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Met Met Ala  
 165 170 175  
 Asp Gly Glu Gln Ile His Ser Ala Met Tyr Pro Gly Ser Ile Phe Gly



```
<400> 160
Met Ser Ser Thr Val Thr Val Ala Ile Ile Gln Ala Ala Pro Val Tyr
 1          5          10          15
Tyr Asp Leu Pro Ala Thr Leu Asp Lys Ala Ala Lys Leu Val Ala Asp
      20          25          30
Ala Ala Ala Gln Gly Ala Thr Leu Ile Val Phe Gly Glu Thr Trp Phe
      35          40          45
```



Pro Gly Tyr Pro Ala Trp Leu Asp Tyr Cys Pro Asn Val Ala Leu Trp  
 50 55 60  
 Asn His Pro Pro Thr Lys Gln Val Phe Glu Arg Leu His Arg Asn Ser  
 65 70 75 80  
 Ile Ala Val Pro Ser Lys Glu Leu Asp Phe Leu Gly Ala Leu Ala Arg  
 85 90 95  
 Lys His Gln Val Val Leu Val Leu Ser Ile Asn Glu Arg Val Glu Gln  
 100 105 110  
 Gly Ala Gly His Gly Thr Leu Tyr Asn Thr Leu Leu Thr Ile Asp Ala  
 115 120 125  
 Asp Gly Thr Leu Ala Asn His His Arg Lys Leu Met Pro Thr Tyr Thr  
 130 135 140  
 Glu Arg Met Val Trp Gly Met Gly Asp Gly Val Gly Leu Gln Ala Val  
 145 150 155 160  
 Asp Thr Ala Val Gly Arg Val Gly Gly Leu Ile Cys Trp Glu His Trp  
 165 170 175  
 Met Pro Leu Ala Arg Gln Thr Met His Ile Ser Gly Glu Gln Ile His  
 180 185 190  
 Ile Ser Val Phe Pro Thr Val His Glu Met His Gln Ile Ala Ser Arg  
 195 200 205  
 Gln Tyr Ala Phe Glu Gly Arg Thr Phe Val Leu Thr Val Gly Gly Ile  
 210 215 220  
 Leu Ala Ala Gln Asp Leu Pro Ala Glu Leu Glu Arg Pro Ala Asp Leu  
 225 230 235 240  
 Pro Pro Thr Gln Leu Val Gln Arg Gly Gly Ser Ala Ile Ile Ala Pro  
 245 250 255  
 Asp Gly Arg Tyr Leu Ala Gly Pro Val Tyr Asn Glu Glu Thr Ile Leu  
 260 265 270  
 Thr Ala Thr Leu Asp Leu Gly Glu Ile Ile Arg Glu Ser Met Thr Leu  
 275 280 285  
 Asp Val Thr Gly His Tyr Ala Arg Pro Asp Val Phe Asp Leu Thr Val  
 290 295 300  
 Lys Arg Ser Arg Pro  
 305

<210> 161  
 <211> 1008  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 161  
 atgaccacca tccgcgcccgc cgccgtgcag tttagcccgg tgctgtactc gcgccaggcc 60  
 accgtcgaca agctgtgccg caccctgctg gaactggggc gcgaaggggt gcagttcgcg 120  
 gtattcccgg aaaccgtggt gccgtactac ccatattttt ccttcgtgca gccaccgttc 180  
 gccatgggca aacaacacct gttgctgctc gagcaatccg tcaactgtgcc ctctgacgtc 240  
 acccggcaga tcggtgaggg ctgccgggaa gcggggatcg tcgccagcat cggcgtcaac 300  
 gaacgcgacg gcggcactat ttataacgcg cagttgctgt tcgatgccga cggcagcctg 360  
 attcagcagc ggcgcaagat caccocgacc tatcacgaac gcatggtctg ggggcagggc 420  
 gatggttccg gcctgcgcgc cgtggacagt gcggtggggc gtatcgggtc cctggcctgc 480  
 tgggaacatt acaaccccct ggcgcgctac gcgctgatgg ccgatggcga acagattcat 540  
 gtggcgatgt ttcccggctc cctggtcggc gacatctttg ccgagcagat cgaagtcacc 600  
 atccgccacc acgccctgga aagcggctgc ttcgtggtca acgccacggc ttggctggat 660  
 gccgaccagc agggccggat catgcaggac accggtgcg agttggggcc gatttccggc 720  
 ggctgtttta ccgcgatcat ttcccggag ggcaagggtc tcggcgagcc gctgcgcagc 780  
 ggcgaagggg tggtcattgc tgacctcgac ctggccctga tcgacaagcg caaacgcatg 840  
 atggattcgg tcggtcacta cagccgcccg gaactgctca gcctgcttat cgaccgcagc 900  
 ccgaccgccc acgtgcatga acttgccgcc gcgcttaatc ctgccaggga gtctgatcca 960  
 ctagtgtcga cctgcaggcg cgcgagctcc agcttttgtt cccttttag 1008



WO 03/097810

PCT/US03/15712

<210> 162  
 <211> 335  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 162  
 Met Thr Thr Ile Arg Ala Ala Ala Val Gln Phe Ser Pro Val Leu Tyr  
 1 5 10 15  
 Ser Arg Gln Ala Thr Val Asp Lys Leu Cys Arg Thr Leu Leu Glu Leu  
 20 25 30  
 Gly Arg Glu Gly Val Gln Phe Ala Val Phe Pro Glu Thr Val Val Pro  
 35 40 45  
 Tyr Tyr Pro Tyr Phe Ser Phe Val Gln Pro Pro Phe Ala Met Gly Lys  
 50 55 60  
 Gln His Leu Leu Leu Leu Glu Gln Ser Val Thr Val Pro Ser Asp Val  
 65 70 75 80  
 Thr Arg Gln Ile Gly Glu Ala Cys Arg Glu Ala Gly Ile Val Ala Ser  
 85 90 95  
 Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Ile Tyr Asn Ala Gln Leu  
 100 105 110  
 Leu Phe Asp Ala Asp Gly Ser Leu Ile Gln Gln Arg Arg Lys Ile Thr  
 115 120 125  
 Pro Thr Tyr His Glu Arg Met Val Trp Gly Gln Gly Asp Gly Ser Gly  
 130 135 140  
 Leu Arg Ala Val Asp Ser Ala Val Gly Arg Ile Gly Ser Leu Ala Cys  
 145 150 155 160  
 Trp Glu His Tyr Asn Pro Leu Ala Arg Tyr Ala Leu Met Ala Asp Gly  
 165 170 175  
 Glu Gln Ile His Val Ala Met Phe Pro Gly Ser Leu Val Gly Asp Ile  
 180 185 190  
 Phe Ala Glu Gln Ile Glu Val Thr Ile Arg His His Ala Leu Glu Ser  
 195 200 205  
 Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala Asp Gln Gln  
 210 215 220  
 Gly Arg Ile Met Gln Asp Thr Gly Cys Glu Leu Gly Pro Ile Ser Gly  
 225 230 235 240  
 Gly Cys Phe Thr Ala Ile Ile Ser Pro Glu Gly Lys Val Leu Gly Glu  
 245 250 255  
 Pro Leu Arg Ser Gly Glu Gly Val Val Ile Ala Asp Leu Asp Leu Ala  
 260 265 270  
 Leu Ile Asp Lys Arg Lys Arg Met Met Asp Ser Val Gly His Tyr Ser  
 275 280 285  
 Arg Pro Glu Leu Leu Ser Leu Leu Ile Asp Arg Ser Pro Thr Ala His  
 290 295 300  
 Val His Glu Leu Ala Ala Ala Leu Asn Pro Ala Arg Glu Ser Asp Pro  
 305 310 315 320  
 Leu Val Ser Thr Cys Arg Arg Ala Ser Ser Phe Cys Ser Leu  
 325 330 335

<210> 163  
 <211> 978  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample



&lt;400&gt; 163

```

gtgaccatca tcaaagccgc cgcagtgcag atcagccccg tgctttacag cccggaagcc      60
accgtcgaaa aggtcgttcg cgagaccgcg gaactcggcc agaagggcgt gcagttcgca      120
acgtttccgg aaaccgtggt gccgtactac ccatacttct ccgccgtcca gacgggcatc      180
gaactgctgt cccgcaaaga gcacctgcga ctgctggagc aggccgtgac tgttccttcc      240
cccgccactg atgcgattgc ccaggcggca cgcgaggccg gcatgggtgt gtcgatcggc      300
gtcaacgagc gtgacggcgg caccatctac aacacgcagc tgctctttga tgccgacggc      360
acgctggtgc agcgccgcgg caagatcacg ccgacgcatt tcgagcgcac ggtgtggggc      420
cagggcgacg gttcggggcct gcgcgcagtg gataccaagg tcggccgcac tggccagctg      480
gcctgcttcg agcacaacaa cccgctcgcg cgctacgcaa tgatggccga tggcgagcag      540
atccattcct ccatgtaccc gggctccgcc ttcggcgacg gattcgcgca gcgcatggag      600
atcaacattc gccaacacgc cctggagtcg ggttgcttcg tggatgaatgc caccgcgtgg      660
ctcgacgccg accagcaggc gcagatcatg aaggacacgg gctgcgccat cgggccgatc      720
tctggcgggt gcttcacgac catcgtcacg ccggacggca tgctgatcgg cgaaccctc      780
cgcgagggcg agggcgagat catcgccgac ctcgatttca ccctgatcga ccgccgcaag      840
ctgctgatgg actcggtcgg ccactacaac cgtccggagc tgctgagcct gctgatcgac      900
cgcacaccgg cggcgaactt ccatgagcgc agtacgcac cggccgtcga tgccgccagc      960
ggcctcgaaa tcctctaa

```

&lt;210&gt; 164

&lt;211&gt; 325

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 164

```

Val Thr Ile Ile Lys Ala Ala Ala Val Gln Ile Ser Pro Val Leu Tyr
  1          5          10          15
Ser Arg Glu Ala Thr Val Glu Lys Val Val Arg Glu Thr Arg Glu Leu
          20          25          30
Gly Gln Lys Gly Val Gln Phe Ala Thr Phe Pro Glu Thr Val Val Pro
          35          40          45
Tyr Tyr Pro Tyr Phe Ser Ala Val Gln Thr Gly Ile Glu Leu Leu Ser
          50          55          60
Gly Lys Glu His Leu Arg Leu Leu Glu Gln Ala Val Thr Val Pro Ser
          65          70          75          80
Pro Ala Thr Asp Ala Ile Ala Gln Ala Ala Arg Glu Ala Gly Met Val
          85          90          95
Val Ser Ile Gly Val Asn Glu Arg Asp Gly Gly Thr Ile Tyr Asn Thr
          100          105          110
Gln Leu Leu Phe Asp Ala Asp Gly Thr Leu Val Gln Arg Arg Arg Lys
          115          120          125
Ile Thr Pro Thr His Phe Glu Arg Met Val Trp Gly Gln Gly Asp Gly
          130          135          140
Ser Gly Leu Arg Ala Val Asp Thr Lys Val Gly Arg Ile Gly Gln Leu
          145          150          155          160
Ala Cys Phe Glu His Asn Asn Pro Leu Ala Arg Tyr Ala Met Met Ala
          165          170          175
Asp Gly Glu Gln Ile His Ser Ser Met Tyr Pro Gly Ser Ala Phe Gly
          180          185          190
Asp Gly Phe Ala Gln Arg Met Glu Ile Asn Ile Arg Gln His Ala Leu
          195          200          205
Glu Ser Gly Cys Phe Val Val Asn Ala Thr Ala Trp Leu Asp Ala Asp
          210          215          220
Gln Gln Ala Gln Ile Met Lys Asp Thr Gly Cys Ala Ile Gly Pro Ile
          225          230          235          240
Ser Gly Gly Cys Phe Thr Thr Ile Val Thr Pro Asp Gly Met Leu Ile
          245          250          255
Gly Glu Pro Leu Arg Glu Gly Glu Gly Glu Ile Ile Ala Asp Leu Asp

```



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
|     |     |     | 260 |     |     |     |     | 265 |     |     |     | 270 |     |     |     |  |  |
| Phe | Thr | Leu | Ile | Asp | Arg | Arg | Lys | Leu | Leu | Met | Asp | Ser | Val | Gly | His |  |  |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |  |  |
| Tyr | Asn | Arg | Pro | Glu | Leu | Leu | Ser | Leu | Leu | Ile | Asp | Arg | Thr | Pro | Ala |  |  |
|     | 290 |     |     |     |     |     | 295 |     |     |     | 300 |     |     |     |     |  |  |
| Ala | Asn | Phe | His | Glu | Arg | Ser | Thr | His | Pro | Ala | Val | Asp | Ala | Ala | Ser |  |  |
| 305 |     |     |     |     | 310 |     |     |     | 315 |     |     |     |     |     | 320 |  |  |
| Gly | Leu | Glu | Ile | Leu |     |     |     |     |     |     |     |     |     |     |     |  |  |
|     |     |     |     | 325 |     |     |     |     |     |     |     |     |     |     |     |  |  |

&lt;210&gt; 165

&lt;211&gt; 1008

&lt;212&gt; DNA

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 165

|            |            |             |            |             |            |      |
|------------|------------|-------------|------------|-------------|------------|------|
| atggccaatt | tcaaattcaa | ggcggcgggcg | gtgcaggccg | cgcccgccttt | cctcgatctc | 60   |
| gaggctagca | tcgccaagtc | gatcgccctg  | atcgaacaag | ccgcccga    | cggcgccaag | 120  |
| ctgatcgctt | ttcccgaagt | cttcattccc  | ggctaccctt | ggcacatctg  | gctcgacagt | 180  |
| ccgcctggg  | cgatcgggcg | cggcttcgtc  | tcgcgctatt | tcgagaactc  | gctggactac | 240  |
| aacagccccg | aggccgagcg | cctcaggctc  | gccgtcaaga | aggcgggctt  | gacggcggtg | 300  |
| atcggcctct | ccgagcgcg  | cggcggcagc  | ctctacatcg | cgcaatggat  | catcggccct | 360  |
| gacggcgaga | ccgttgcgaa | acggcgtaag  | ctccggccga | cccattgcga  | gcgcacggtc | 420  |
| tatggagaag | gcgacggcag | cgacctcgcg  | gttcacgacg | tatctggcat  | cggccgtctc | 480  |
| ggcgcgctct | gctgctggga | gcatatccag  | ccgctgtcga | aattcgcgat  | gtattcgcaa | 540  |
| aatgagcaag | tgcacgtcgc | gtcctggccg  | agcttctcgc | tctacgaccc  | gttcgcgcgc | 600  |
| gcgctgggcg | ccgaggtcaa | caacgcagcc  | tcgcggatct | atgcggtcga  | aggctcatgc | 660  |
| ttcgatcatt | cgccctgcgc | gaccgtttcg  | cctgcaatga | tcgaggaact  | gtgcgacgcg | 720  |
| ccaaacaaac | atgcgcttct | gcacgcgggc  | ggcggcttcg | cgcgcatcta  | tgggccggac | 780  |
| ggcgcttcga | tcgccgagac | gctgccgcca  | gatcaggaag | gcttgatcta  | cgccgacatc | 840  |
| gacctcaccg | cgatcggcgt | cgccaaggcc  | gccgccgatc | ccgccggcca  | ttattcgcgc | 900  |
| cccagcgta  | cgcgctgtct | cttcaacaag  | aagcccgtct | ggcgagtcga  | aacttttgct | 960  |
| ttgcccgtcg | atgcgccggc | gccggagacg  | cagaccgccg | cgagctga    |            | 1008 |

&lt;210&gt; 166

&lt;211&gt; 335

&lt;212&gt; PRT

&lt;213&gt; Unknown

&lt;220&gt;

&lt;223&gt; Obtained from an environmental sample

&lt;400&gt; 166

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
| Met | Ala | Asn | Phe | Lys | Phe | Lys | Ala | Ala | Ala | Val | Gln | Ala | Ala | Pro | Ala |  |  |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     |     | 15  |     |  |  |
| Phe | Leu | Asp | Leu | Glu | Ala | Ser | Ile | Ala | Lys | Ser | Ile | Ala | Leu | Ile | Glu |  |  |
|     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |  |  |
| Gln | Ala | Ala | Ala | Asn | Gly | Ala | Lys | Leu | Ile | Ala | Phe | Pro | Glu | Val | Phe |  |  |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |  |  |
| Ile | Pro | Gly | Tyr | Pro | Trp | His | Ile | Trp | Leu | Asp | Ser | Pro | Ala | Trp | Ala |  |  |
|     | 50  |     |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |  |  |
| Ile | Gly | Arg | Gly | Phe | Val | Ser | Arg | Tyr | Phe | Glu | Asn | Ser | Leu | Asp | Tyr |  |  |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |  |  |
| Asn | Ser | Pro | Glu | Ala | Glu | Arg | Leu | Arg | Leu | Ala | Val | Lys | Lys | Ala | Gly |  |  |
|     |     |     | 85  |     |     |     | 90  |     |     |     |     |     |     | 95  |     |  |  |
| Leu | Thr | Ala | Val | Ile | Gly | Leu | Ser | Glu | Arg | Asp | Gly | Gly | Ser | Leu | Tyr |  |  |
|     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |     |  |  |
| Ile | Ala | Gln | Trp | Ile | Ile | Gly | Pro | Asp | Gly | Glu | Thr | Val | Ala | Lys | Arg |  |  |



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | 115 |     | 120 |     | 125 |     |     |     |     |     |     |     |     |     |     |
| Arg | Lys | Leu | Arg | Pro | Thr | His | Cys | Glu | Arg | Thr | Val | Tyr | Gly | Glu | Gly |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Asp | Gly | Ser | Asp | Leu | Ala | Val | His | Asp | Val | Ser | Gly | Ile | Gly | Arg | Leu |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Gly | Ala | Leu | Cys | Cys | Trp | Glu | His | Ile | Gln | Pro | Leu | Ser | Lys | Phe | Ala |
|     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Met | Tyr | Ser | Gln | Asn | Glu | Gln | Val | His | Val | Ala | Ser | Trp | Pro | Ser | Phe |
|     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Ser | Leu | Tyr | Asp | Pro | Phe | Ala | Pro | Ala | Leu | Gly | Ala | Glu | Val | Asn | Asn |
|     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Ala | Ala | Ser | Arg | Ile | Tyr | Ala | Val | Glu | Gly | Ser | Cys | Phe | Val | Ile | Ala |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Pro | Cys | Ala | Thr | Val | Ser | Pro | Ala | Met | Ile | Glu | Glu | Leu | Cys | Asp | Ala |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Pro | Asn | Lys | His | Ala | Leu | Leu | His | Ala | Gly | Gly | Gly | Phe | Ala | Arg | Ile |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Tyr | Gly | Pro | Asp | Gly | Ala | Ser | Ile | Ala | Glu | Thr | Leu | Pro | Pro | Asp | Gln |
|     |     | 260 |     |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Glu | Gly | Leu | Ile | Tyr | Ala | Asp | Ile | Asp | Leu | Thr | Ala | Ile | Gly | Val | Ala |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Lys | Ala | Ala | Ala | Asp | Pro | Ala | Gly | His | Tyr | Ser | Arg | Pro | Asp | Val | Thr |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Arg | Leu | Leu | Phe | Asn | Lys | Lys | Pro | Ala | Arg | Arg | Val | Glu | Thr | Phe | Ala |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Leu | Pro | Val | Asp | Ala | Pro | Ala | Pro | Glu | Thr | Gln | Thr | Ala | Ala | Ser |     |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |

<210> 167  
 <211> 1017  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Obtained from an environmental sample

<400> 167

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| atgggtattg | aacatccgaa | gtacagggtt | gccgtggtgc | aggccgcacc | ggcctggctc | 60   |
| gatcttgacg | cgtcgatcga | caagtcgatc | gcgctgatcg | aggaggctgc | ccagaaaggc | 120  |
| gccaagctga | tcgcattccc | cgaggccttc | atccccggct | acccttgcca | tatctggatg | 180  |
| gactcgcccg | cctgggcgat | tggccgcggt | tttgtgcagc | gctacttcga | caattcgctg | 240  |
| gcctatgaca | gcccgcaggc | cgagaagctg | cgcgcgcccg | tgcgcaaggc | aaaactcacg | 300  |
| gccgtgatcg | gcttgctcga | gcgtgacggc | ggcagccttt | atctcgacac | atggctgatc | 360  |
| ggccccgacg | gcgagaccat | cgcaaaacgg | cgcaagctgc | ggccgacaca | tgccgagcgc | 420  |
| actgtgtacg | gcgagggcga | cggcagcgac | cttgcggtcc | acaatcgtcc | ggacatcggc | 480  |
| aggctcggtg | cgctctgctg | ctgggagcat | cttcagccac | tgtcgaaata | cgcgatgtac | 540  |
| gcgcagaacg | agcaggtgca | cgtcgcggcc | tggccgagct | tttcgctcta | cgatcccttc | 600  |
| gccgtggcgc | tcggcgccga | ggtgaacaac | gcggcctccc | gcgtctatgc | ggtcgaaggc | 660  |
| tcctgcttcg | tgctggcgcc | gtgcgcgaca | gtctcgcaag | ccatgatcga | cgagctctgc | 720  |
| gatcgggcgg | acaagcacgc | gctgctgcat | gtcggcgggc | gctttgccc  | gatctacggg | 780  |
| cccgcggcga | gccagatcgg | cgacaagctc | gccccgacc  | aggagggcct | ggtgatcgcc | 840  |
| gagatcgatc | tcggcgccat | aggtgtcgcc | aagaacgccc | cggatcccgc | cgggcactat | 900  |
| tcgcggcccc | acgtgacgcg | gctgttgctc | aacaagaaac | cgtacaagcg | cgtcgaacag | 960  |
| ttctcgccgc | cgtcggaggc | ggttgaaccc | acggatatcg | cggcggcggc | aagctga    | 1017 |

<210> 168  
 <211> 338  
 <212> PRT  
 <213> Unknown

<220>



# **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_2\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

# **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_2\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.



What is claimed is:

1. An isolated or recombinant nucleic acid encoding a polypeptide having nitrilase activity, the nucleic acid comprising: a) a nucleotide having a sequence at least 90% identical to SEQ ID NO: 209; b) a variant of SEQ ID NO: 209 having a mutation selected from: nucleotide positions 163-165 of SEQ ID NO: 209 are substituted with AAA, AAG, GGT, GGC, GGA, GGG, CAA, or CAG; nucleotide positions 178-180 of SEQ ID NO: 209 are substituted with GAA or GAG; nucleotide positions 331-333 of SEQ ID NO: 209 are substituted with TCT, TCC, TCA, TCG, AGT, or AGC; nucleotide positions 568-570 of SEQ ID NO: 209 are substituted with CAT, CAC, TCT, TCC, TCA, TCG, AGT, ACT, ACC, ACA, TCA, TAT, TAC, ATG or ACG; nucleotide positions 571-573 of SEQ ID NO: 209 are substituted with TTA, TTG, CTT, CTC, CTA, CTG, GTT, GTC, GTA, GTG, ATG, ACT, ACC, ACA, GAT, GAC, GGT, GGC, GGA, GGG, GAA, GAG, TAC, or ACG; nucleotide positions 595-597 of SEQ ID NO: 209 are substituted with GAA, GAG, TTA, TTG, CTT, CTC, CTA, or CTG; and nucleotide positions 664-666 of SEQ ID NO: 209 are substituted with TTA, TTG, CTT, CTA, or CTG; or c) a nucleic acid having any combination of the mutations in b).
2. A nucleic acid vector capable of replication in a host cell, wherein the vector comprises the nucleic acid of claim 1.
3. A host cell comprising the nucleic acid of claim 1.
4. An isolated or recombinant nucleic acid encoding a polypeptide having nitrilase activity comprising amino acids having a sequence at least 90% identical to SEQ ID NO: 210, or variants of SEQ ID NO: 210 having one or more mutations: residue 55 of SEQ ID NO: 210 is substituted with lysine, glycine, or glutamine; residue 60 of SEQ ID NO: 210 is substituted with glutamic acid; residue 190 of SEQ ID NO: 210 is substituted with histidine, tyrosine or threonine; residue 191 of SEQ ID NO: 210 is substituted with leucine, valine, methionine, aspartic acid, glycine, glutamic acid, tyrosine or threonine; or residue 199 of SEQ ID NO: 210 is substituted with glutamic acid or leucine.



5. An isolated or recombinant nucleic acid encoding a polypeptide variant of SEQ ID NO: 210 having one or more mutations: residue 55 of SEQ ID NO: 210 is substituted with lysine, glycine, or glutamine; residue 60 of SEQ ID NO: 210 is substituted with glutamic acid; residue 190 of SEQ ID NO: 210 is substituted with histidine, tyrosine or threonine; residue 191 of SEQ ID NO: 210 is substituted with leucine, valine, methionine, aspartic acid, glycine, glutamic acid, tyrosine or threonine; or residue 199 of SEQ ID NO: 210 is substituted with glutamic acid or leucine; wherein the polypeptide variant has nitrilase activity.
6. The nucleic acid of claim 1, wherein the nucleic acid is affixed to a solid support.
7. The nucleic acid of claim 6, wherein the solid support is selected from the group of a gel, a resin, a polymer, a ceramic, a glass, a microelectrode and any combination thereof.
8. An isolated or recombinant polypeptide having nitrilase activity, the polypeptide comprising: a) an amino acid sequence having at least 90% identity to SEQ ID NO: 210; b) an amino acid sequence variant of SEQ ID NO: 210 having a mutation selected from: amino acid residue 55 of SEQ ID NO: 210 is substituted with lysine, glycine, or glutamine; amino acid residue 60 of SEQ ID NO: 210 is substituted with glutamic acid; amino acid residue 190 of SEQ ID NO: 210 is substituted with histidine, tyrosine or threonine; amino acid residue 191 of SEQ ID NO: 210 is substituted with leucine, valine, methionine, aspartic acid, glycine, glutamic acid, tyrosine or threonine; or amino acid residue 199 of SEQ ID NO: 210 is substituted with glutamic acid or leucine; or c) an amino acid sequence of SEQ ID NO: 210 and a mutation at all the amino acid residues in b).
9. The polypeptide of claim 8, wherein the polypeptide is affixed to a solid support.
10. The polypeptide of claim 9, wherein the solid support is selected from the group consisting of a gel, a resin, a polymer, a ceramic, a glass, a microelectrode and any combination thereof.



11. A purified antibody that specifically binds to the polypeptide of claim 8.
12. An enzyme preparation which comprises at least one of the polypeptides of claim 8, wherein the preparation is liquid or dry.
13. The enzyme preparation of claim 12, wherein the preparation is affixed to a solid support.
14. A method for hydrolyzing a cyanohydrin moiety or an aminonitrile moiety of a molecule, the method comprising contacting the molecule with the polypeptide of claim 8.
15. A method for making a chiral alpha-hydroxy acid molecule or a chiral amino acid molecule, the method comprising admixing a molecule having a cyanohydrin moiety or an aminonitrile moiety with the polypeptide of claim 8.
16. A method for making an (*R*)-ethyl 4-cyano-3-hydroxybutyric acid, the method comprising contacting a hydroxyglutaryl nitrile with the polypeptide of claim 8 that selectively produces an (*R*)-enantiomer, so as to make (*R*)-ethyl 4-cyano-3-hydroxybutyric acid.
17. A method for making an (*S*)-ethyl 4-cyano-3-hydroxybutyric acid, the method comprising contacting a hydroxyglutaryl nitrile with the polypeptide of claim 8 that selectively produces an (*S*)-enantiomer, so as to make (*S*)-ethyl 4-cyano-3-hydroxybutyric acid.
18. A method for making an (*R*)-mandelic acid, the method comprising admixing a mandelonitrile with the polypeptide of claim 8.
19. A method for making an (*S*)-mandelic acid, the method comprising admixing a mandelonitrile with the polypeptide of claim 8.
20. A method for making an (*S*)-phenyl lactic acid derivative or an (*R*)-phenyl lactic acid derivative, the method comprising admixing a phenyllactocyanonitrile with the polypeptide of claim 8 that selectively produces an (*S*)-



enantiomer or an (*R*)-enantiomer, thereby producing an (*S*)-phenyl lactic acid derivative or an (*R*)-phenyl lactic acid derivative.

21. A method for making the polypeptide of claim 8, the method comprising:
  - (a) introducing a nucleic acid encoding the polypeptide into a host cell under conditions that permit production of the polypeptide by the host cell, and
  - (b) recovering the polypeptide so produced.
22. A screening assay for identifying a nitrilase, the assay comprising:
  - (a) providing a plurality of nucleic acids or polypeptides comprising at least one of the nucleic acids of any one of claims 1, 4 and 5, or at least one of the polypeptides of claim 8;
  - (b) obtaining polypeptide candidates to be tested for nitrilase activity from the plurality;
  - (c) testing the candidates for nitrilase activity; and
  - (d) identifying those polypeptide candidates which are nitrilases.
23. A kit comprising (a) the nucleic acid of any one of claims 1, 4 and 5, or (b) the polypeptide of claim 8 and (c) a buffer.
24. The isolated or recombinant nucleic acid of any one of claims 1 and 4, having at least 95% sequence identity to SEQ ID NO: 209.
25. The isolated or recombinant nucleic acid of any one of claims 1 and 4, having at least 98% sequence identity to SEQ ID NO: 209.
26. The isolated or recombinant nucleic acid of any one of claims 1 and 4, having at least 99% sequence identity to SEQ ID NO: 209.
27. An isolated or recombinant polypeptide having a nitrilase activity comprising a polypeptide having a sequence with at least 90% sequence identity to SEQ ID NO: 210.
28. The isolated or recombinant polypeptide of claim 27, wherein the polypeptide has at least 95% sequence identity to SEQ ID NO: 210.



29. The isolated or recombinant polypeptide of claim 27, wherein the polypeptide has at least 98% sequence identity to SEQ ID NO: 210.
30. The isolated or recombinant polypeptide of claim 29, wherein the polypeptide has the sequence of SEQ ID NO: 210.
31. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolyzing a nitrile to a carboxylic acid.
32. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolyzing a cyanohydrin moiety or an aminonitrile moiety of a molecule.
33. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises an enantio-selective nitrilase activity and the enantio-selective activity comprises generating a chiral alpha-hydroxy acid molecule or a chiral amino acid molecule by catalyzing the hydrolysis of a molecule having a cyanohydrin moiety or an aminonitrile moiety.
34. The isolated or recombinant polypeptide of claim 33, wherein the generated chiral molecule is an (R)-enantiomer.
35. The isolated or recombinant polypeptide of claim 33, wherein the generated chiral molecule is an (S)-enantiomer.
36. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolysis of a hydroxyglutaryl nitrile to generate an ethyl 4-cyano-3-hydroxybutyric acid.
37. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolysis of a hydroxyglutaryl nitrile to generate an (R)-ethyl 4-cyano-3-hydroxybutyric acid.
38. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolysis of a mandelonitrile to generate a mandelic acid or mandelic acid derivative.



39. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the nitrilase activity comprises hydrolysis of phenyllactocyanonitrile to generate an (S)-phenyl lactic acid derivative or an (R)-phenyl lactic acid derivative.

40. The isolated or recombinant polypeptide of any one of claims 27 to 30, wherein the polypeptide is made by a method comprising introducing a nucleic acid encoding the polypeptide into a host cell, wherein the host cell produces the polypeptide, and optionally the method further comprises recovering the polypeptide.



Figure 1

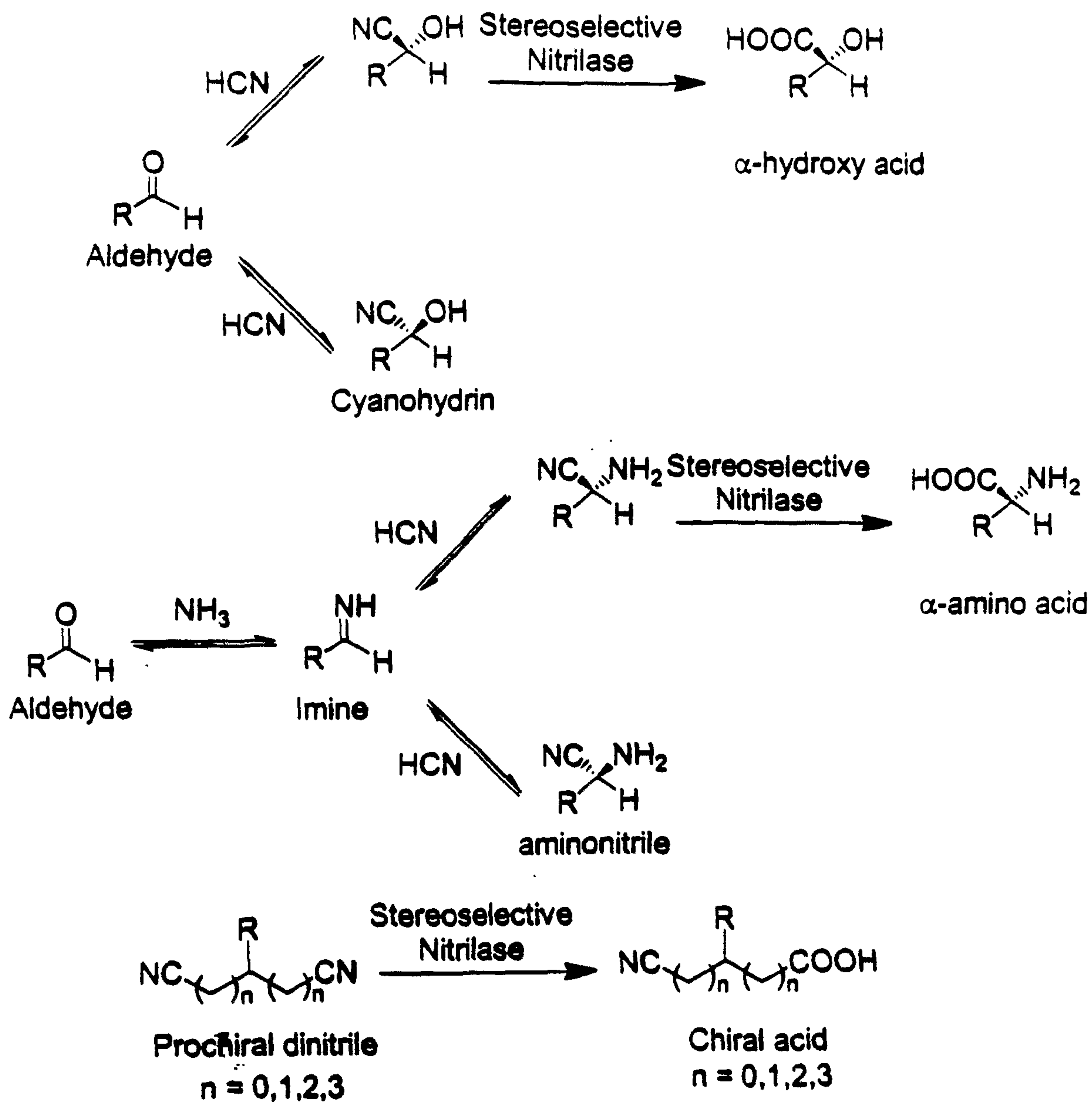




Figure 2

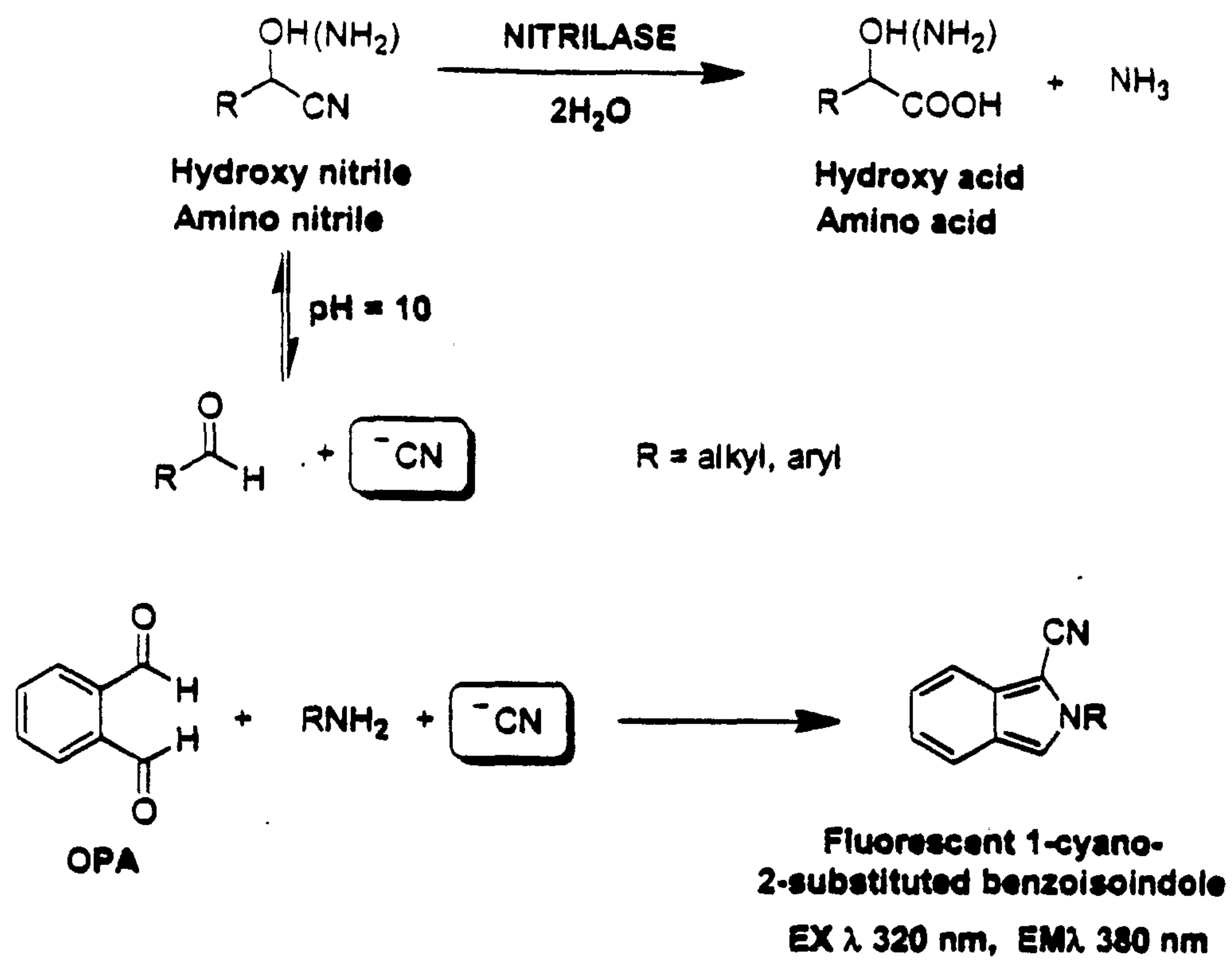




Figure 3

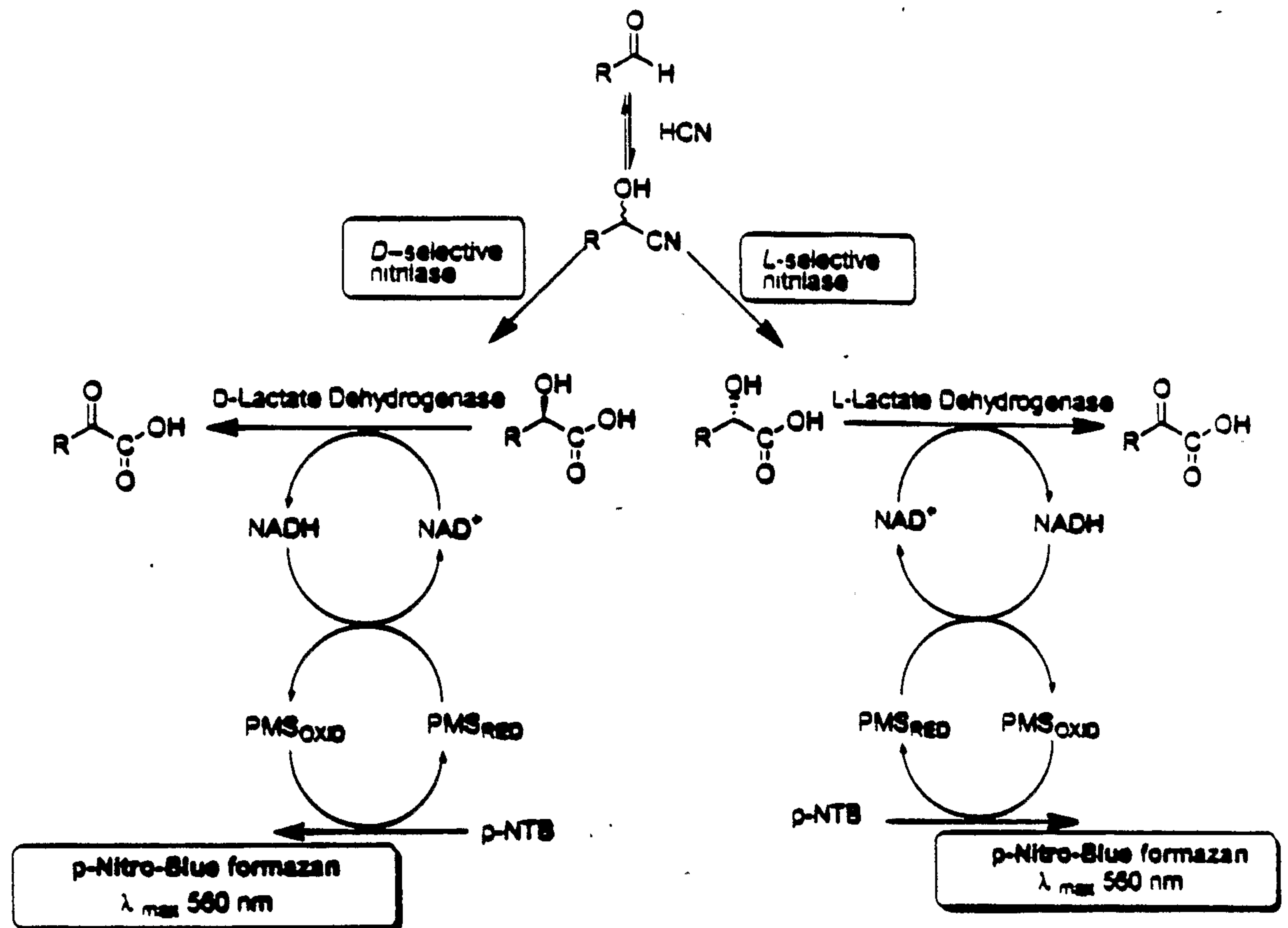
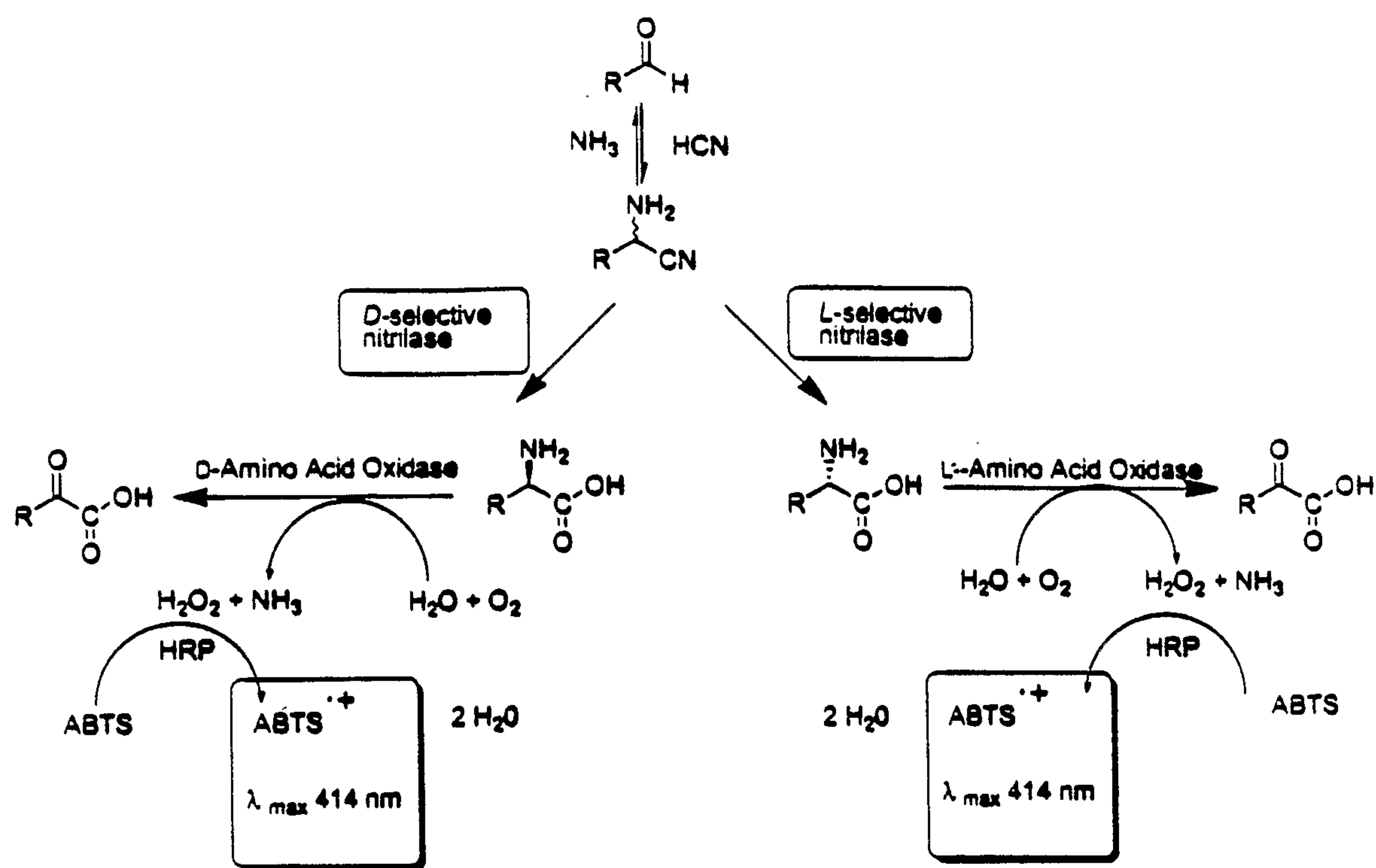




Figure 4





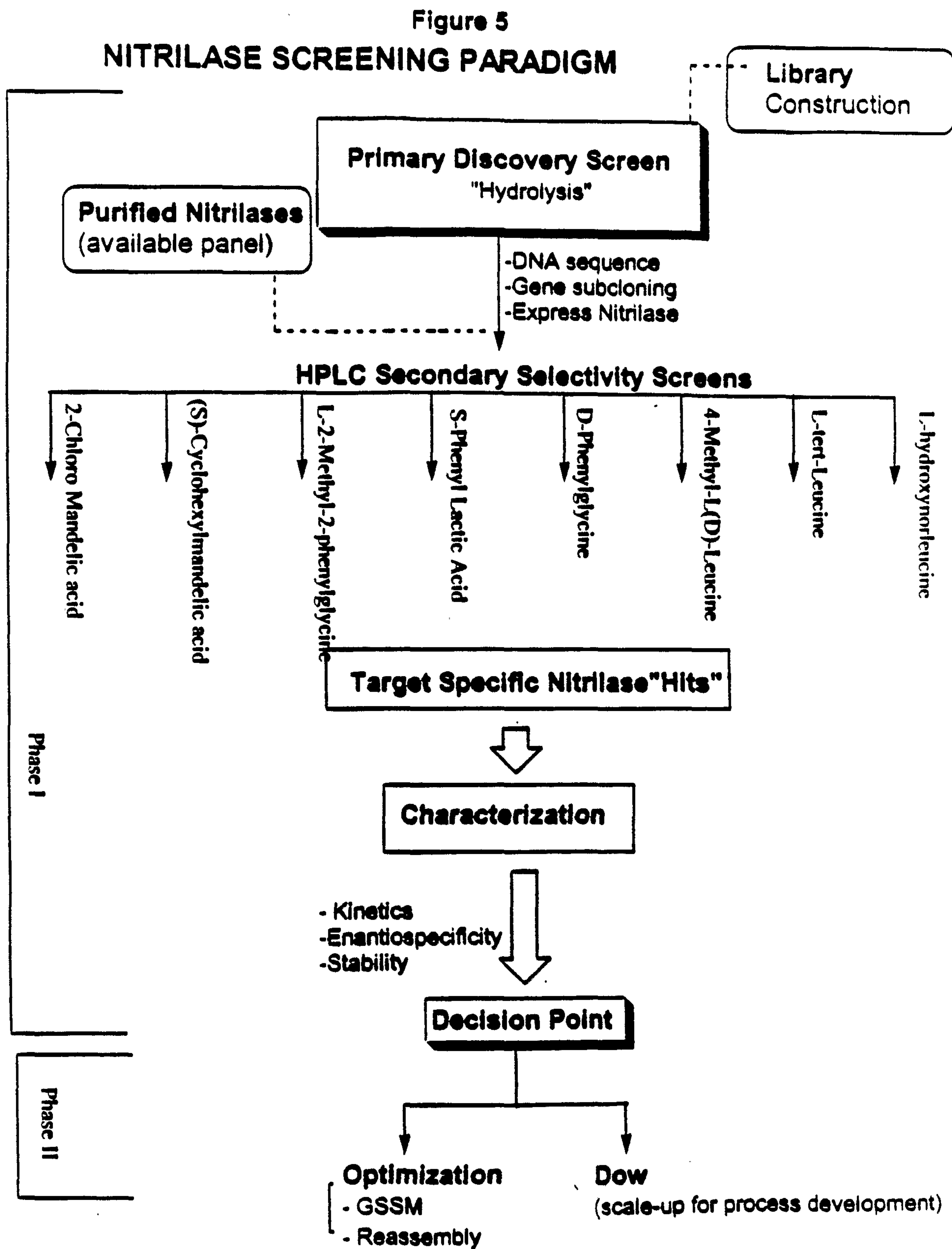




Figure 6A

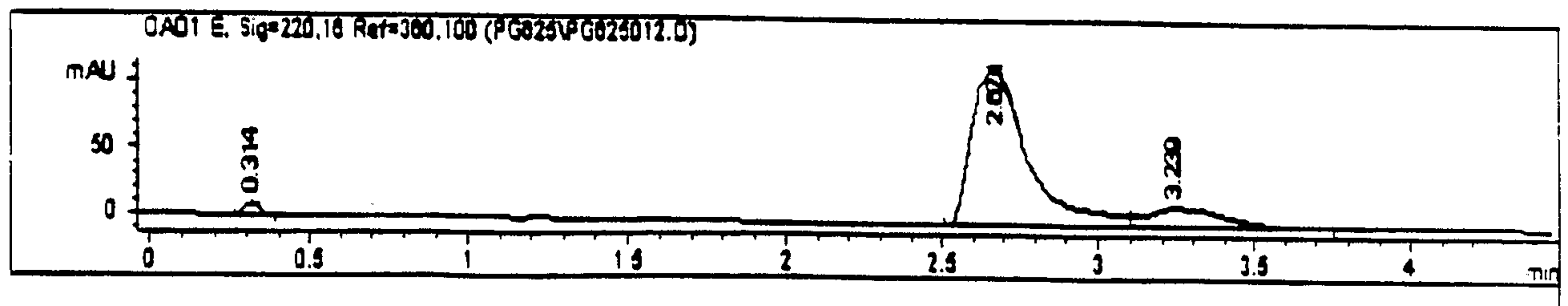


Figure 6B

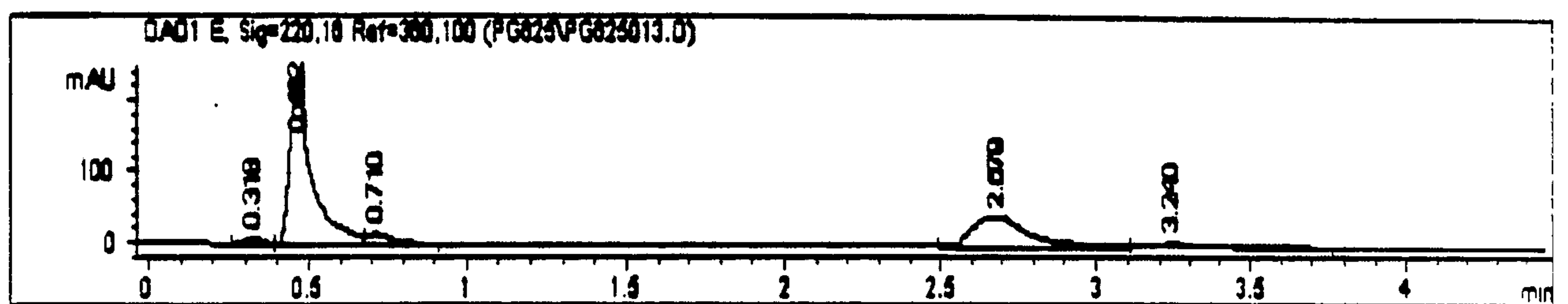
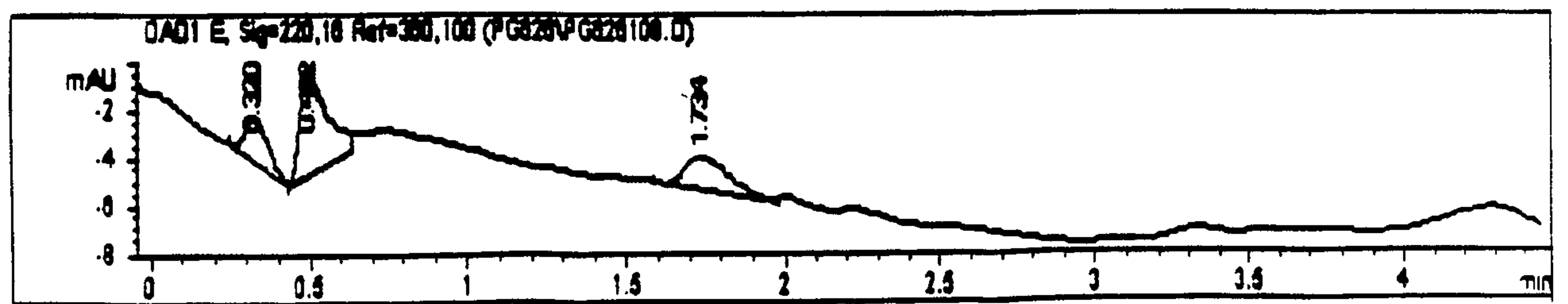
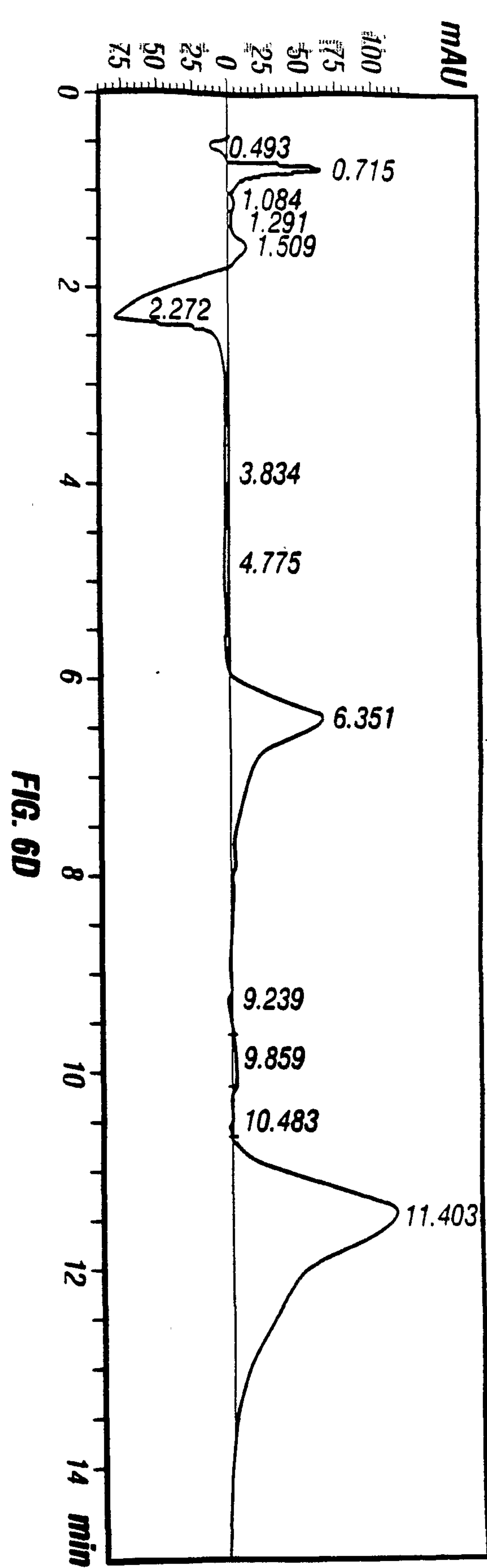
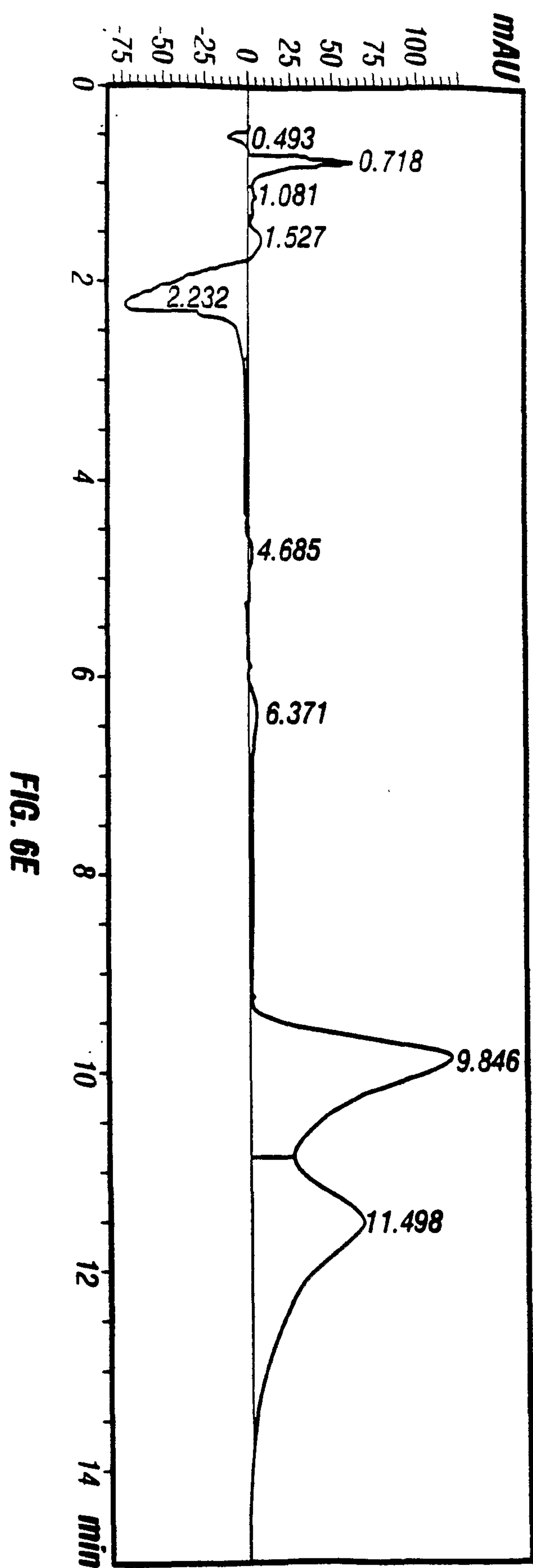


Figure 6C









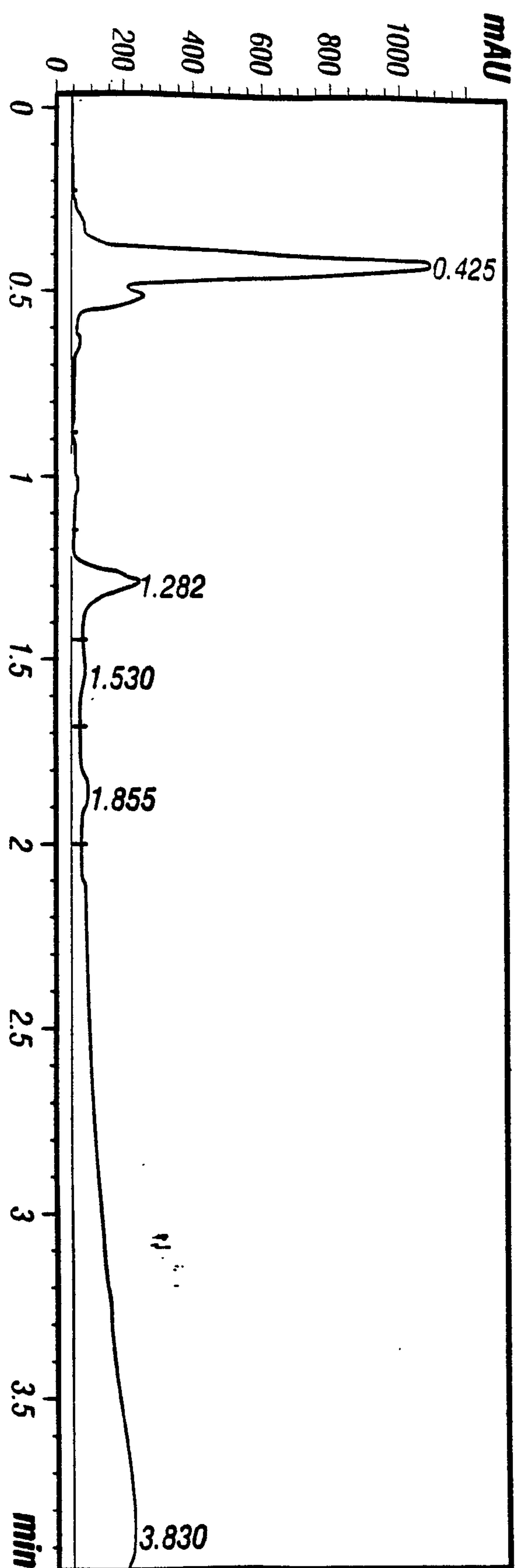


FIG. 7B

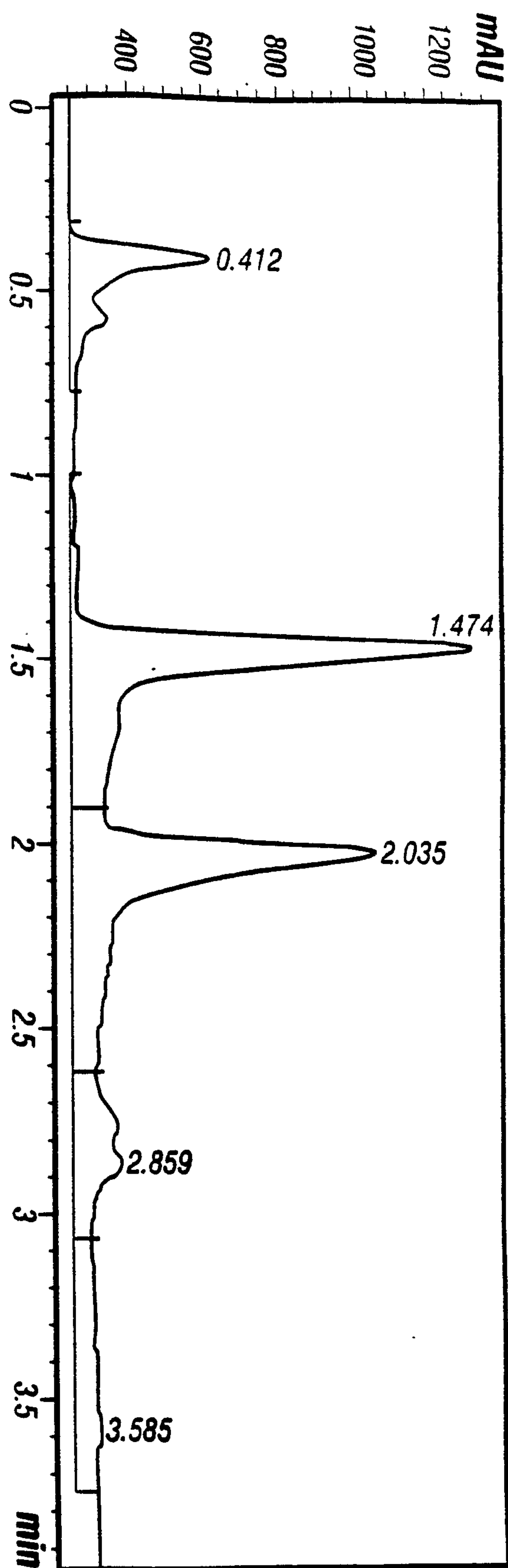
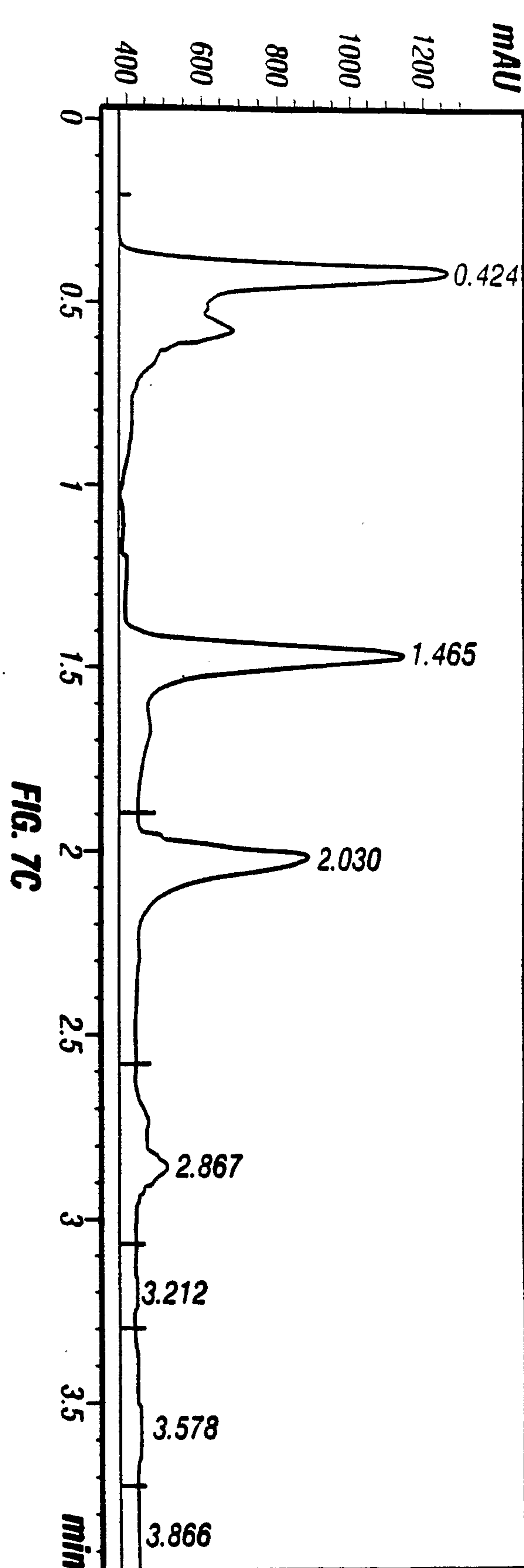
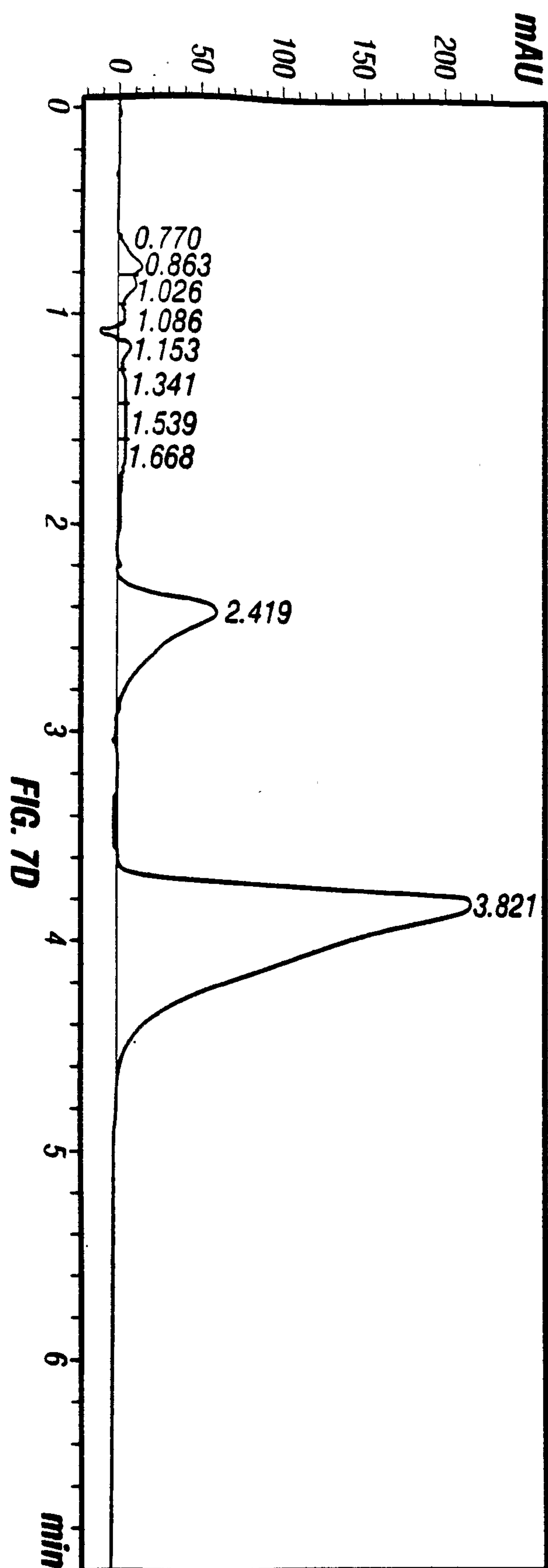
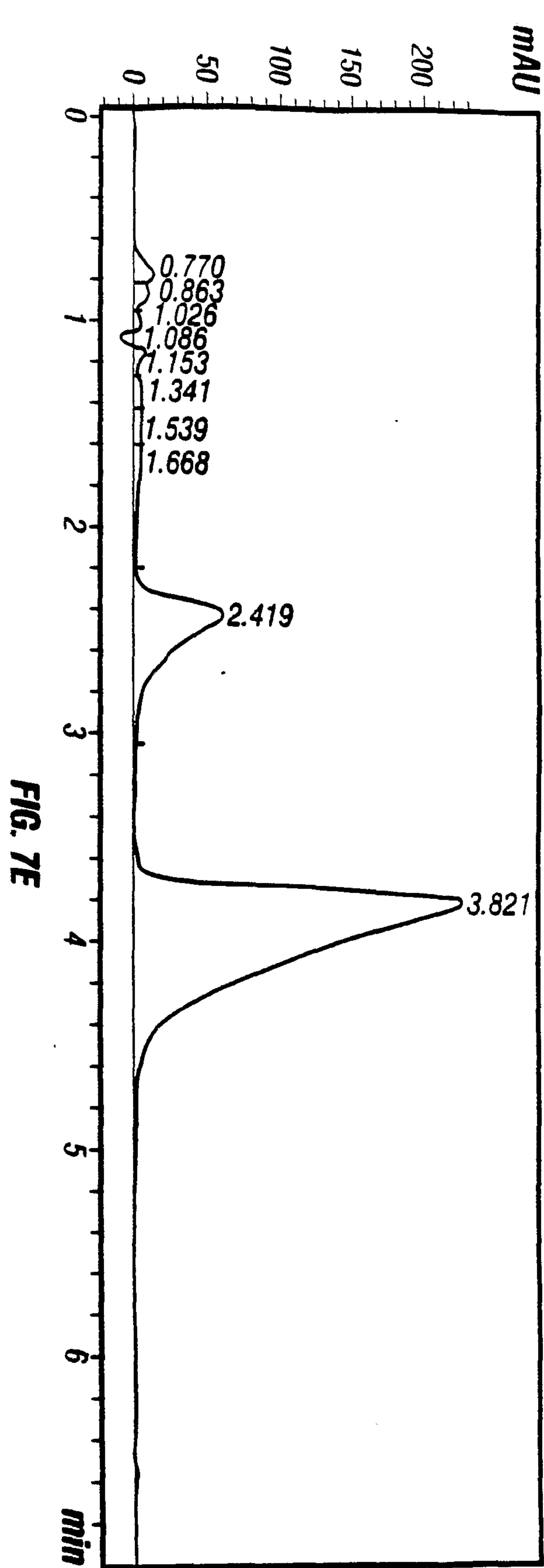
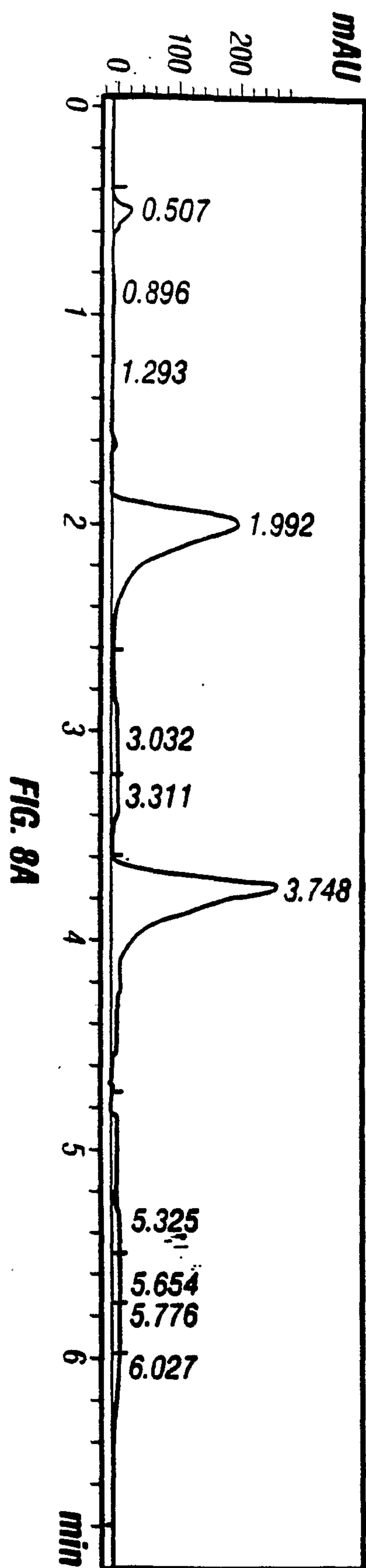


FIG. 7A

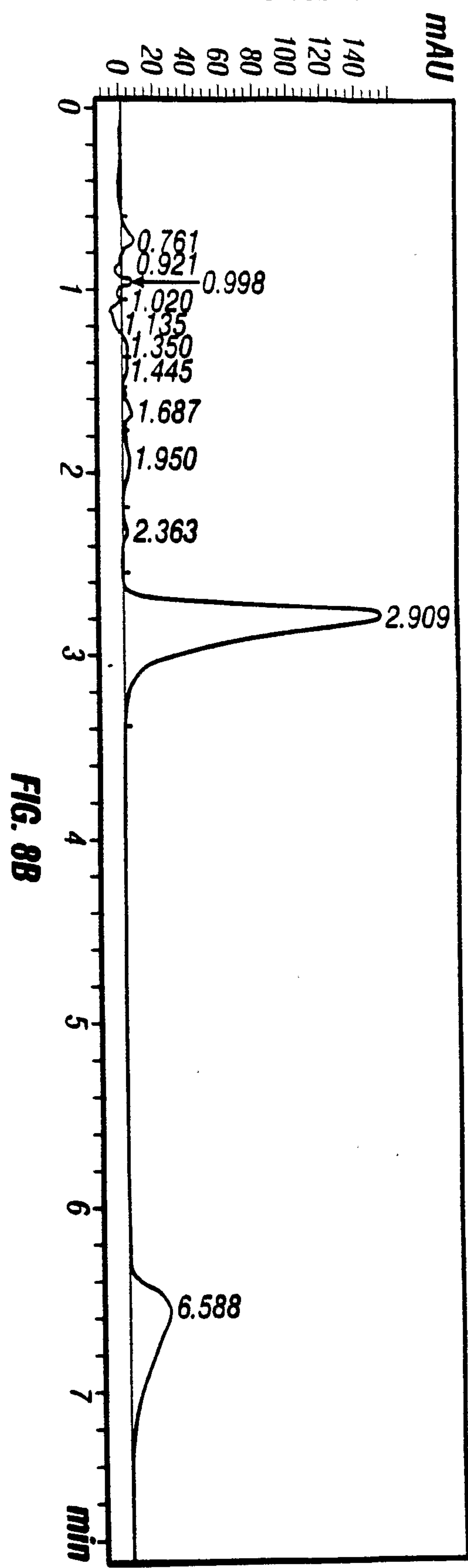
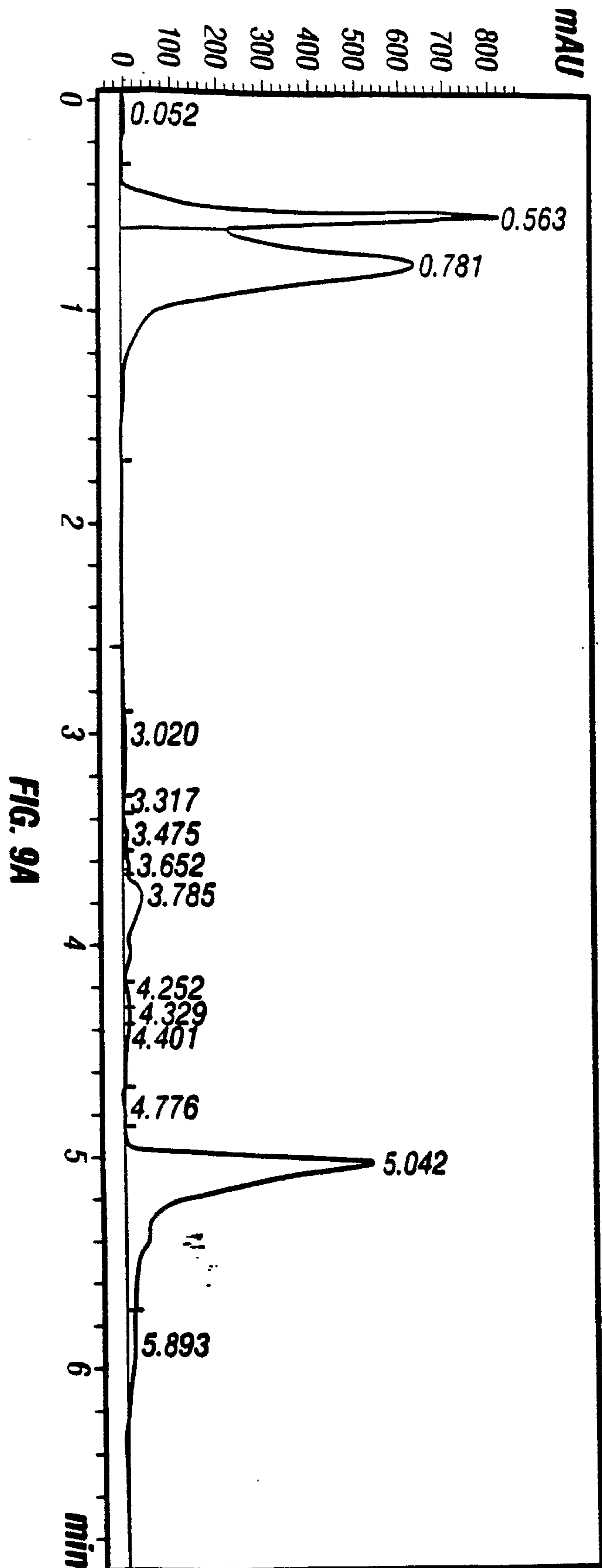




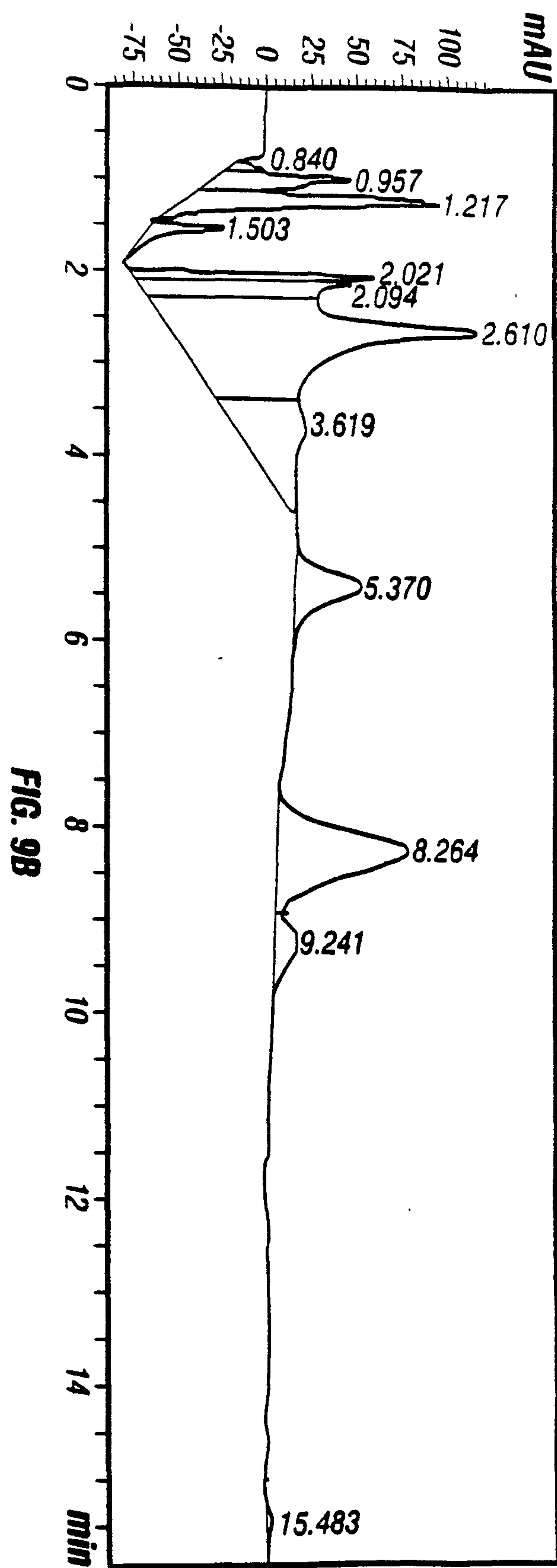
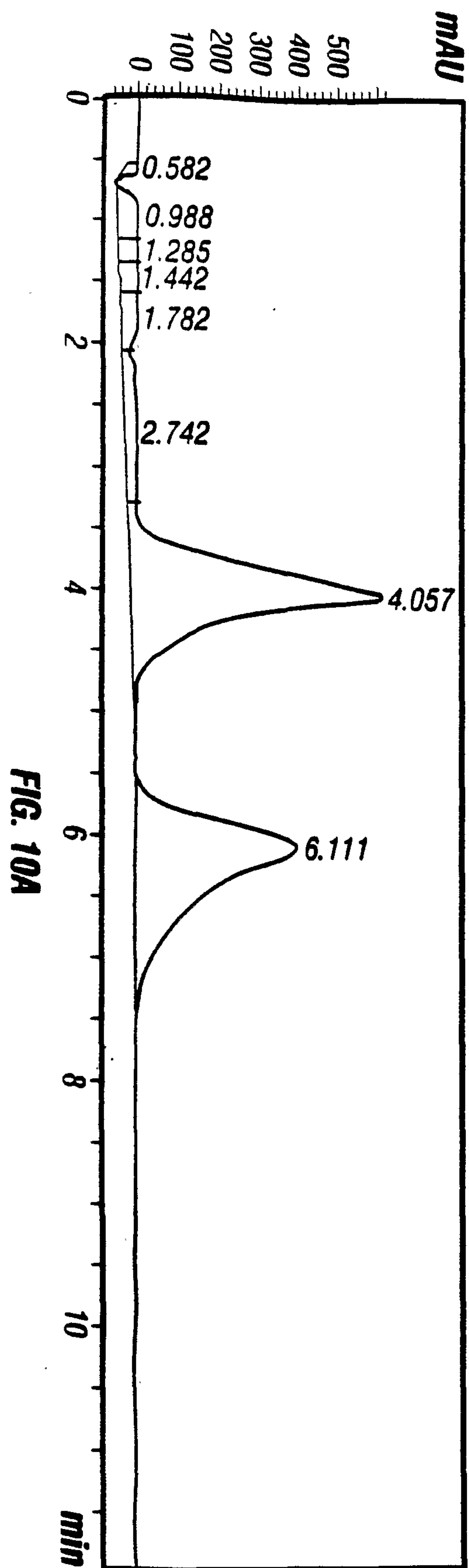




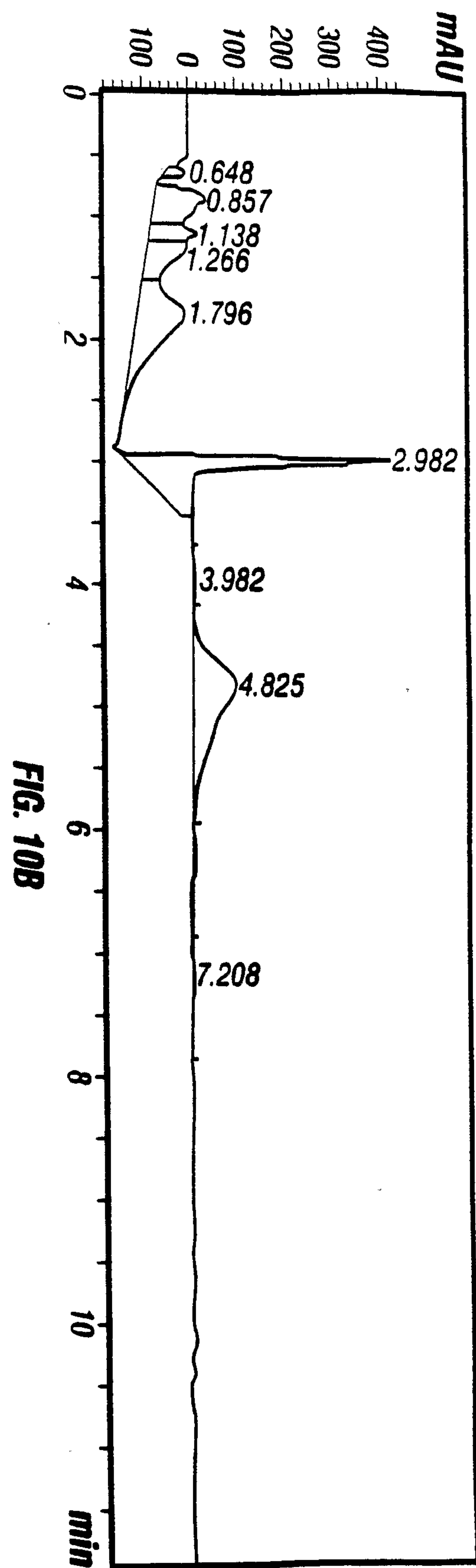
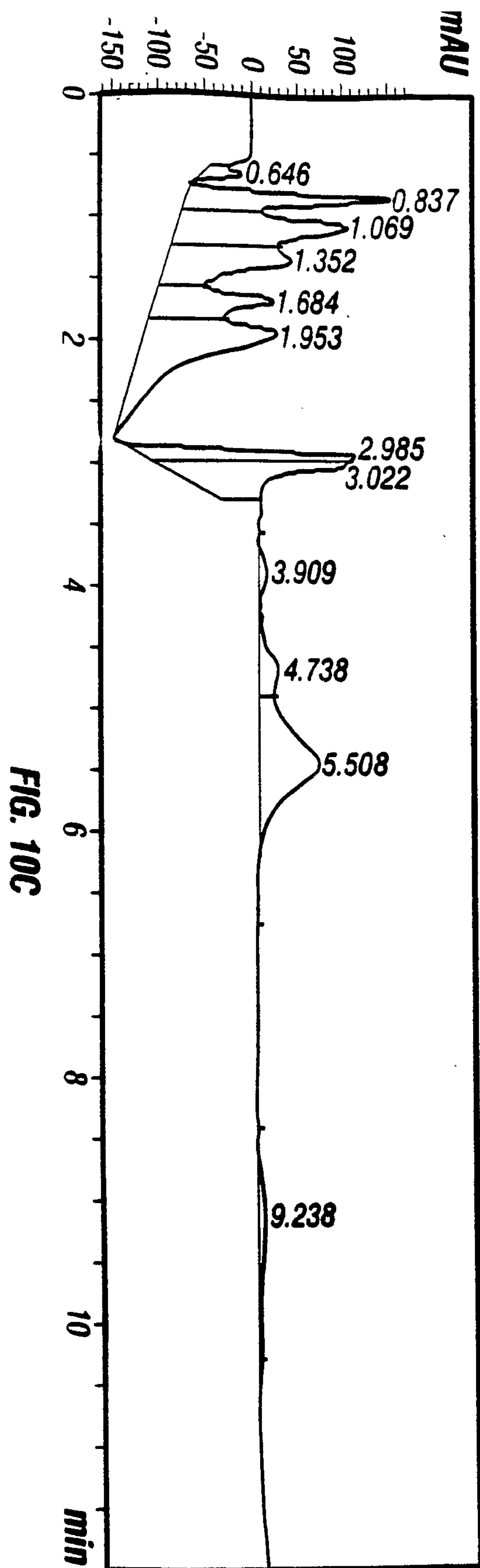




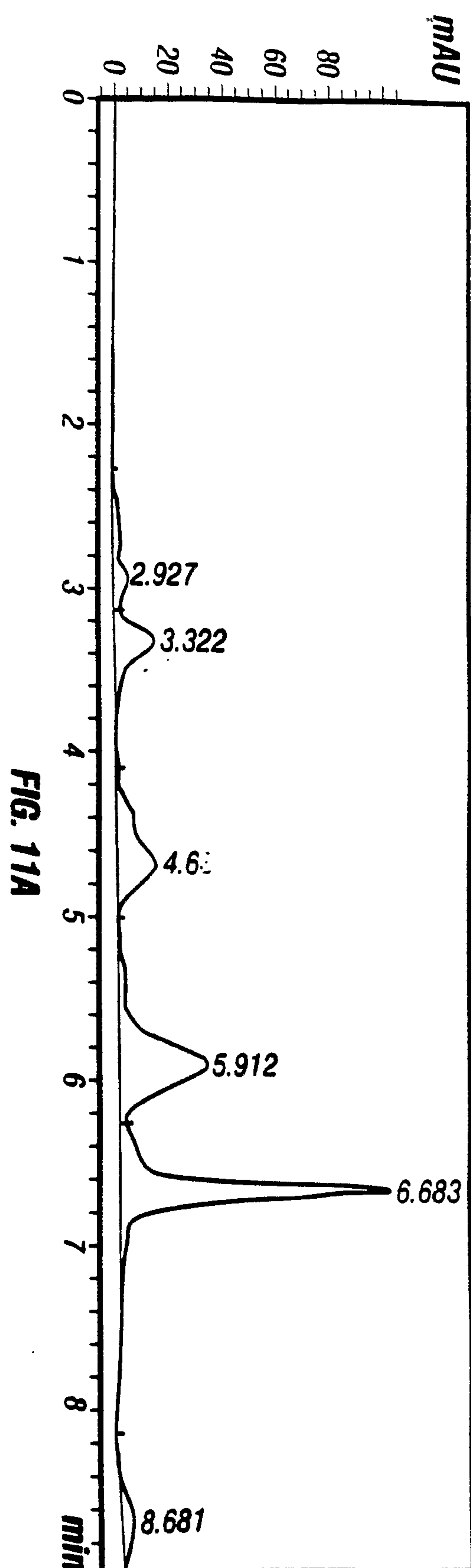
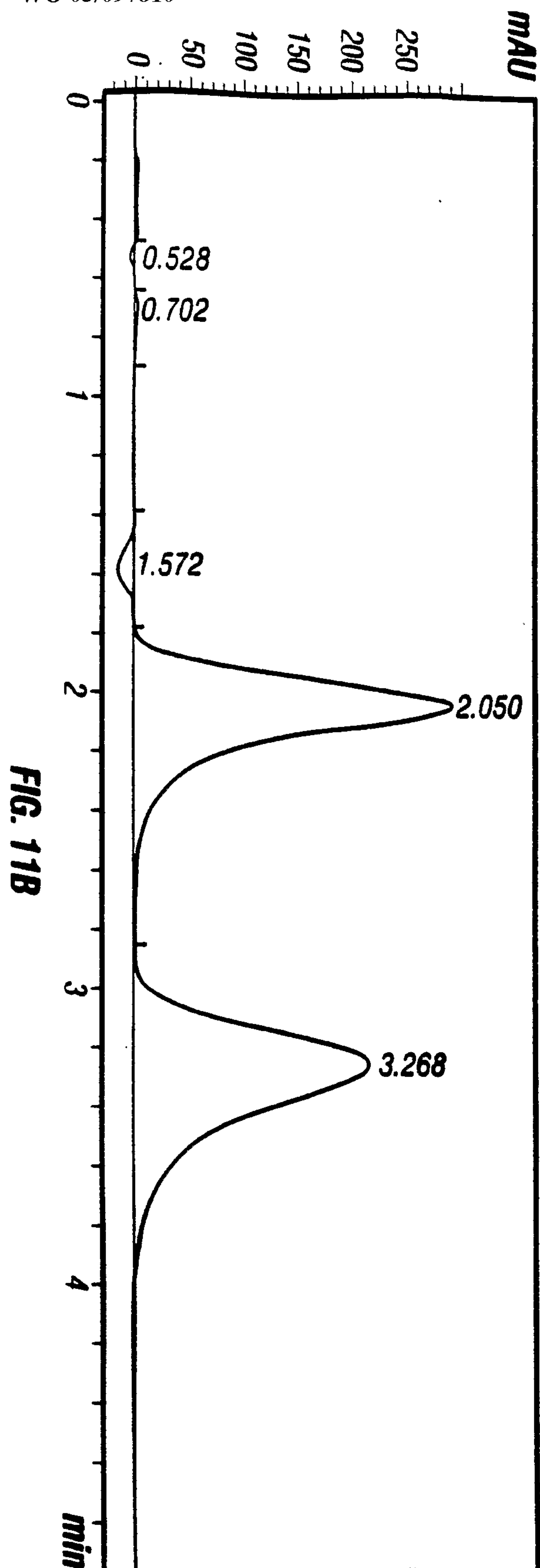




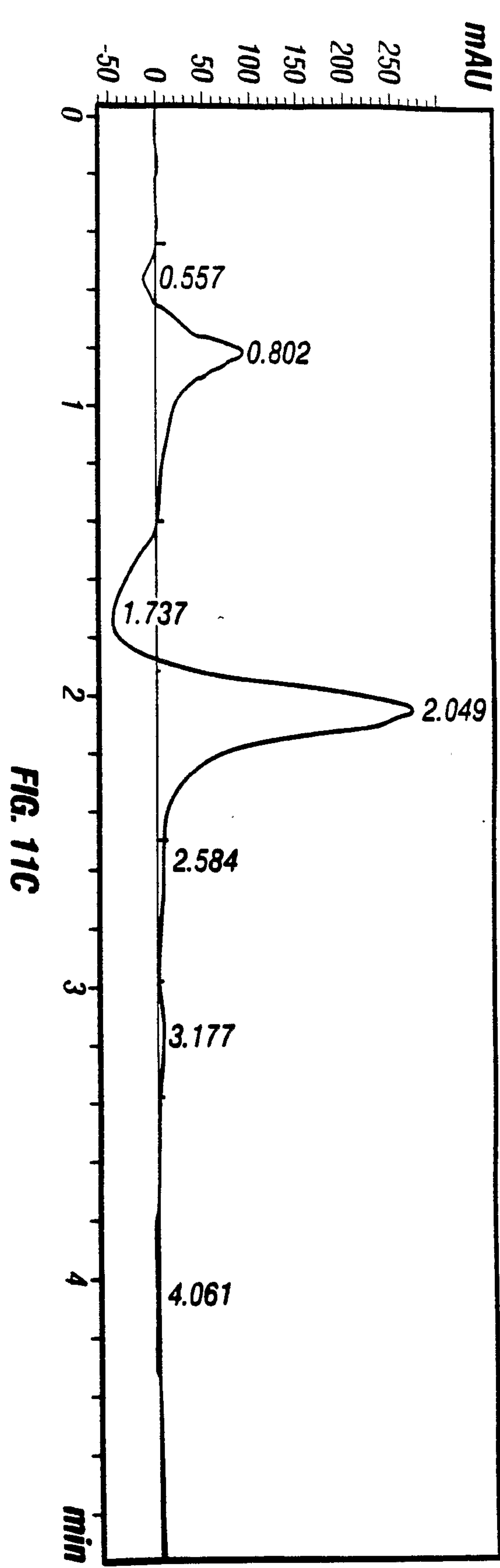
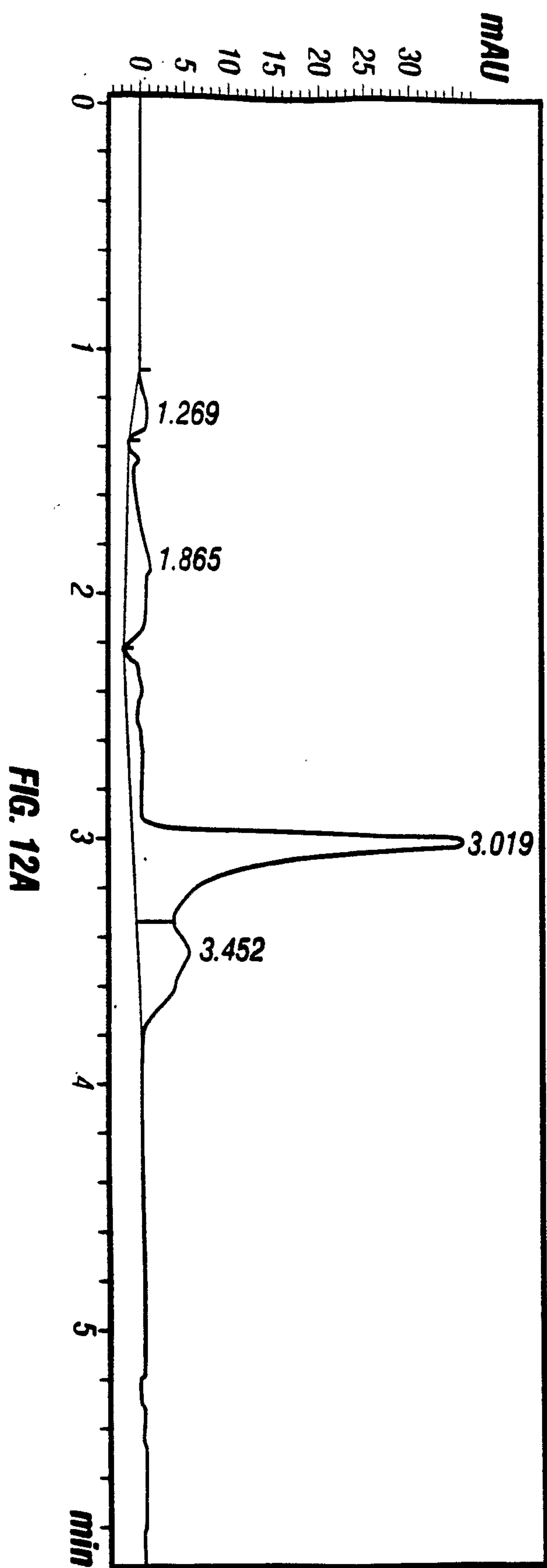




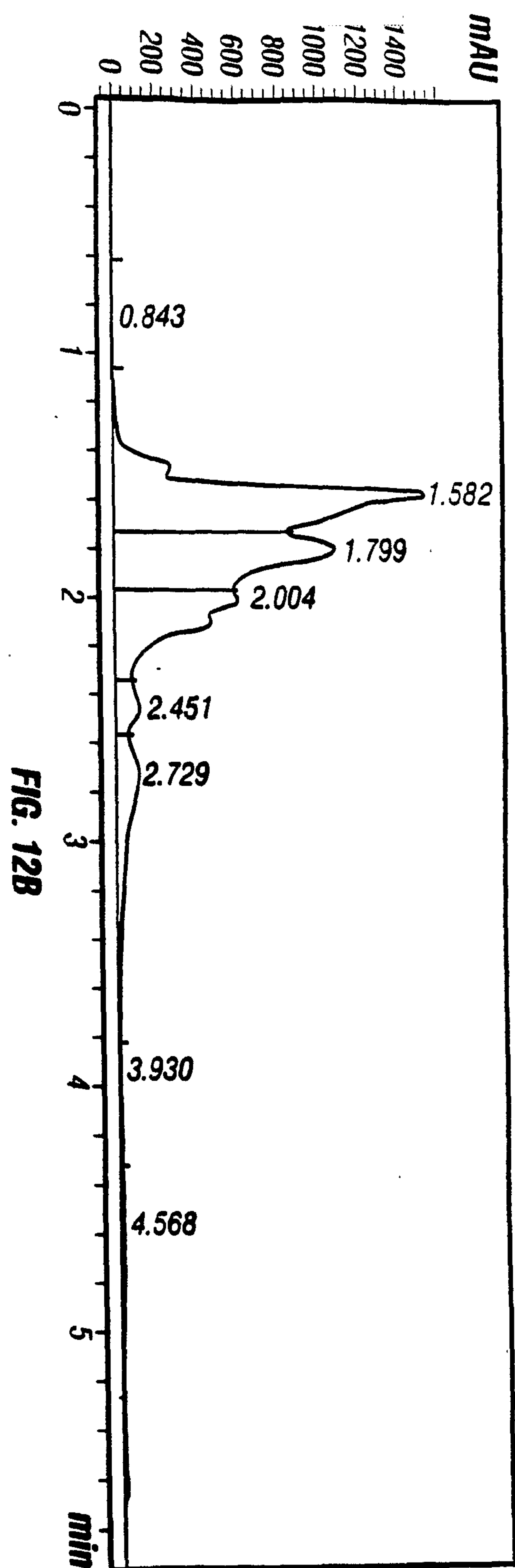
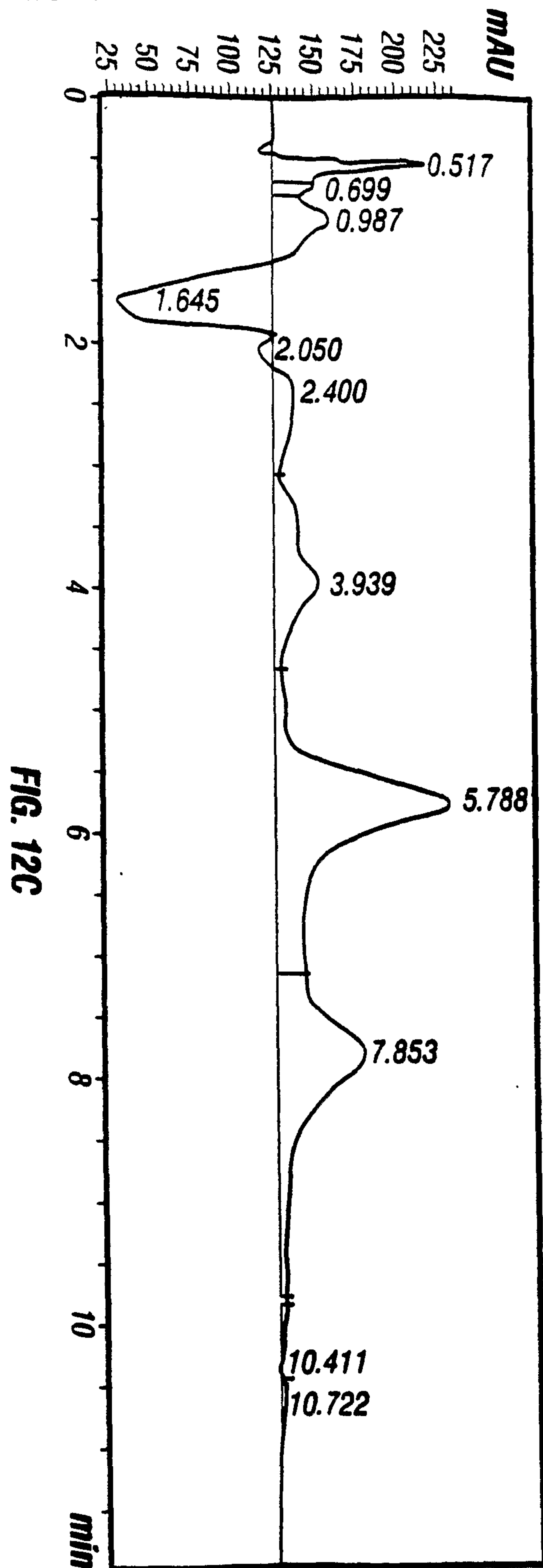














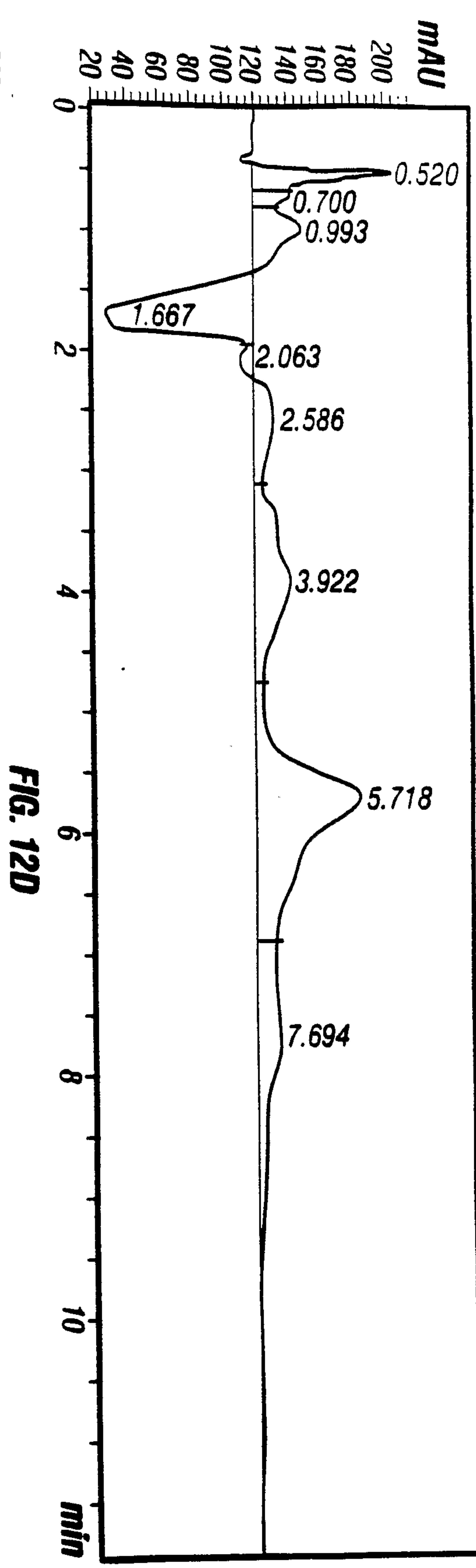
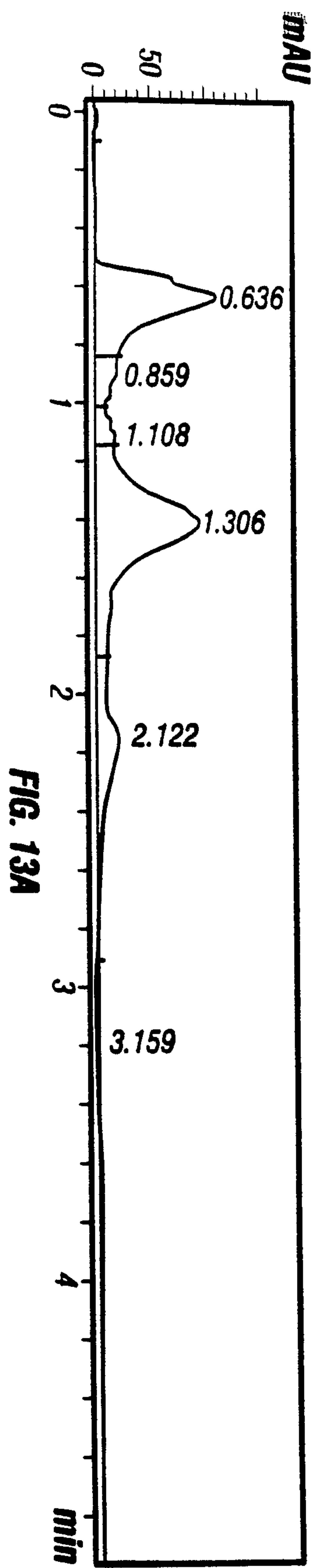
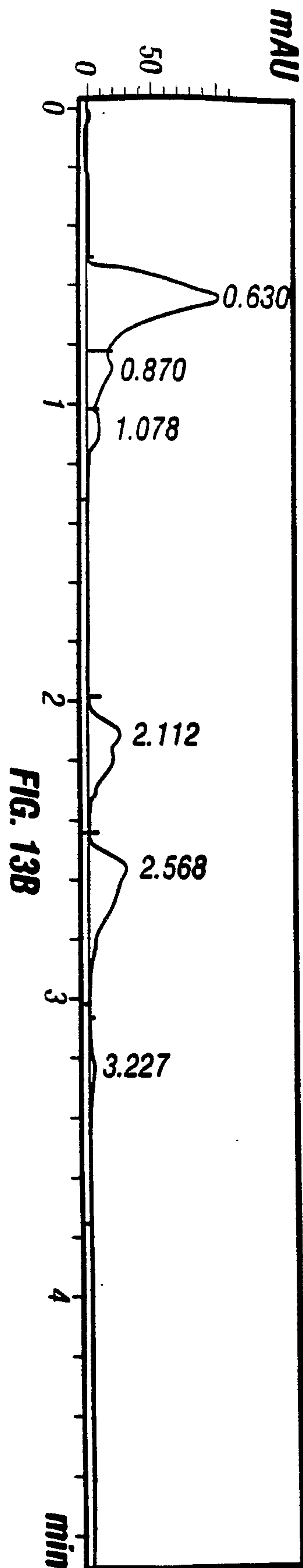




FIGURE 14A

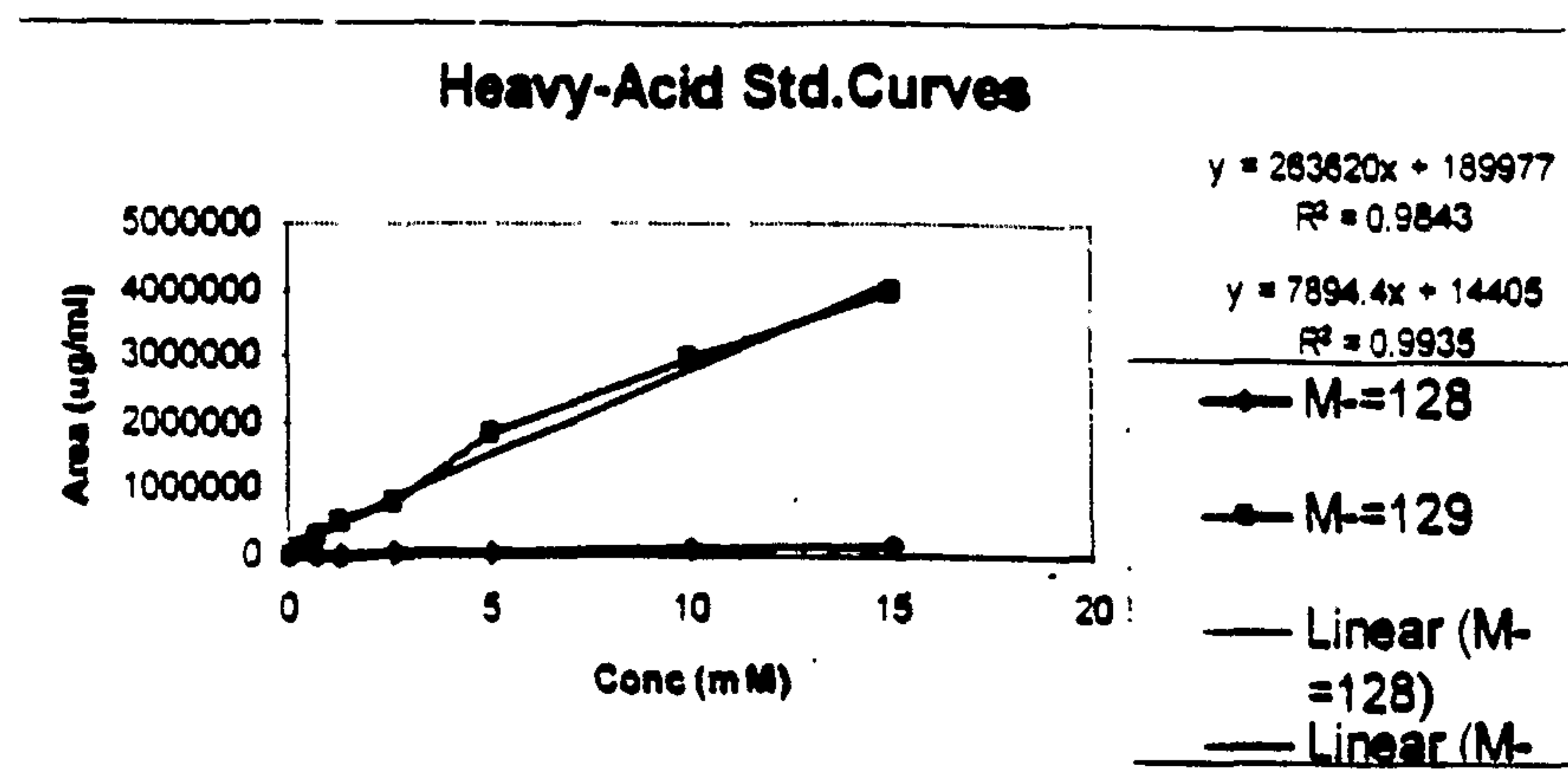


FIGURE 14B

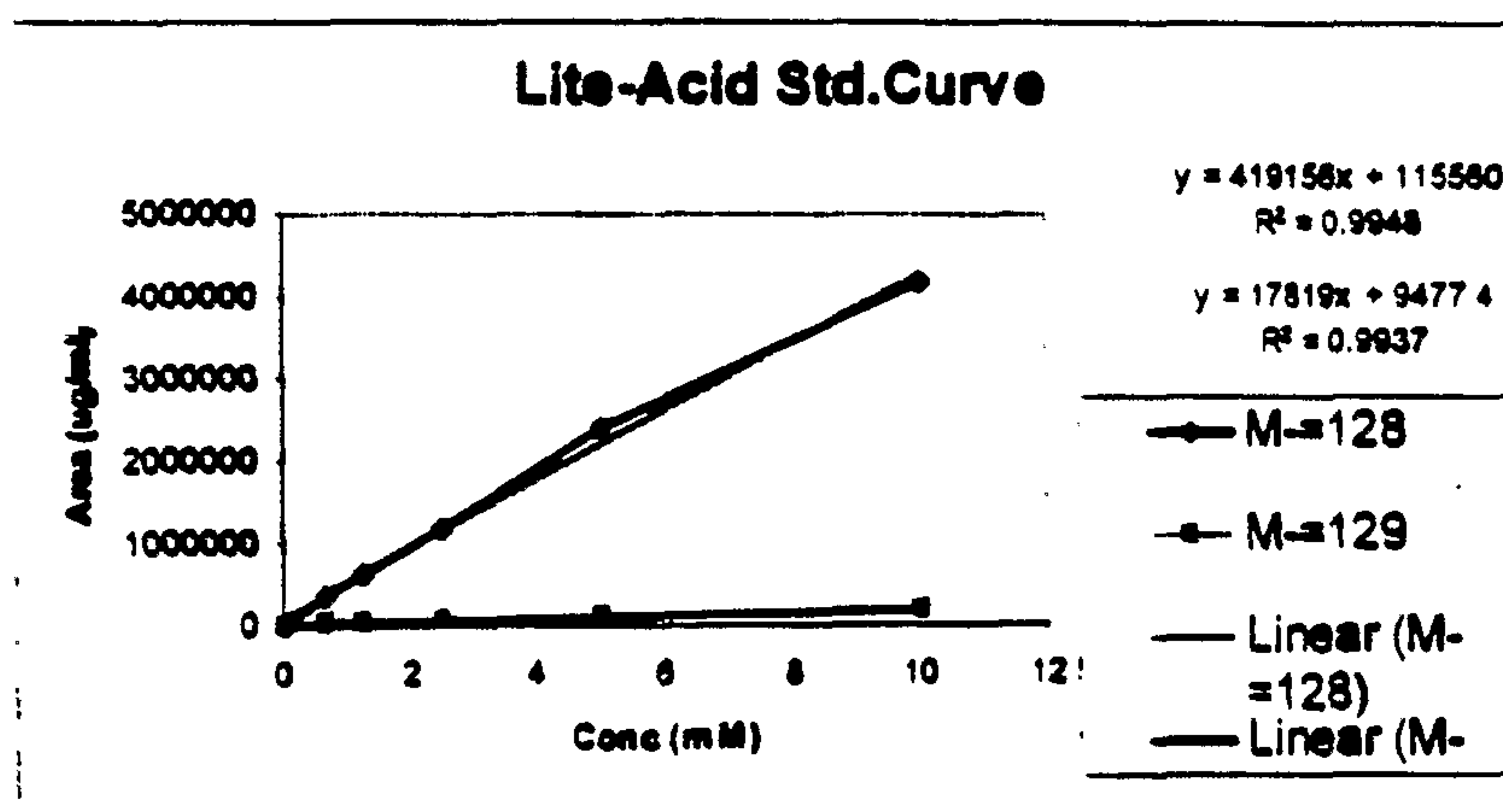




Figure 15

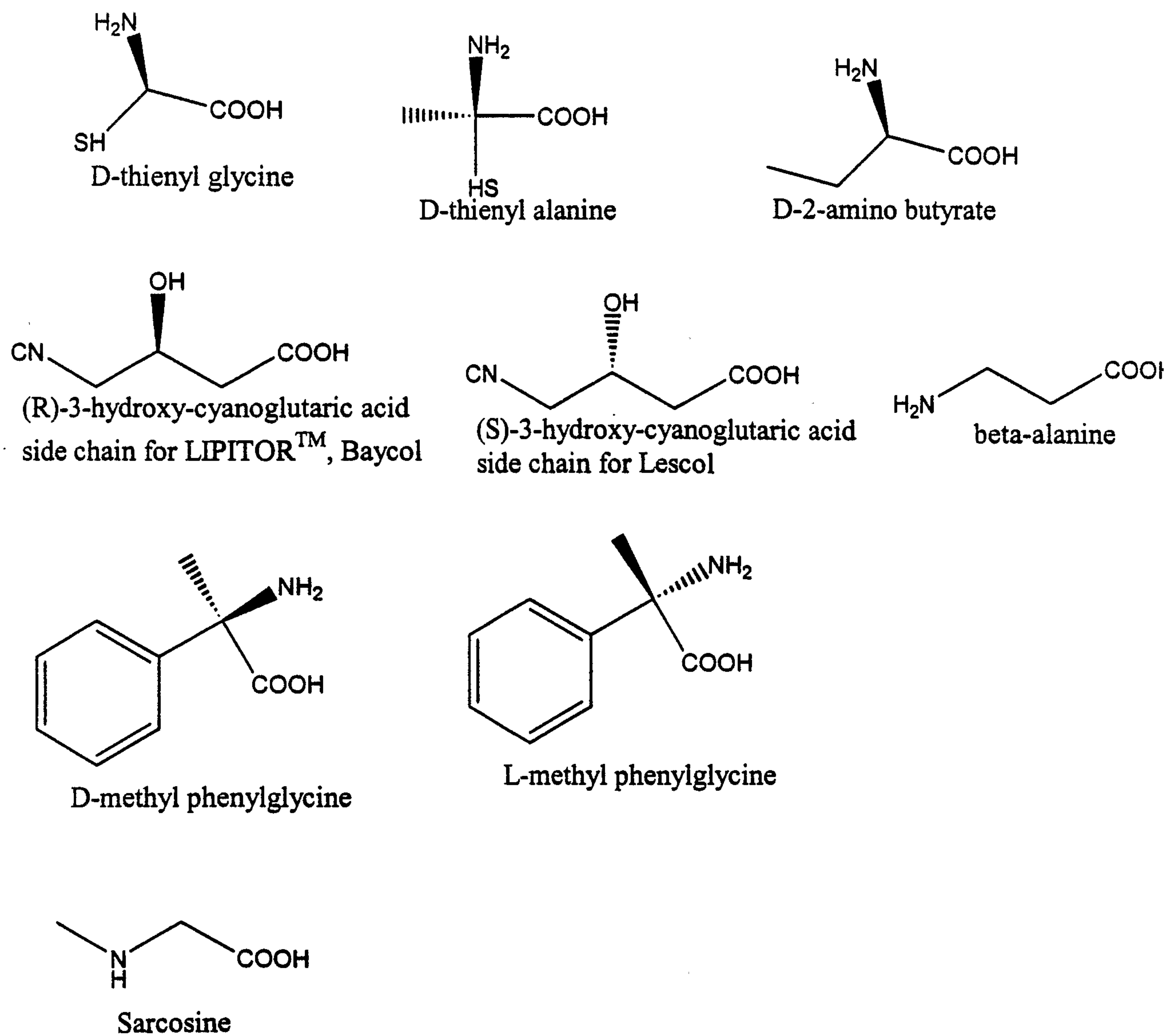




Figure 15

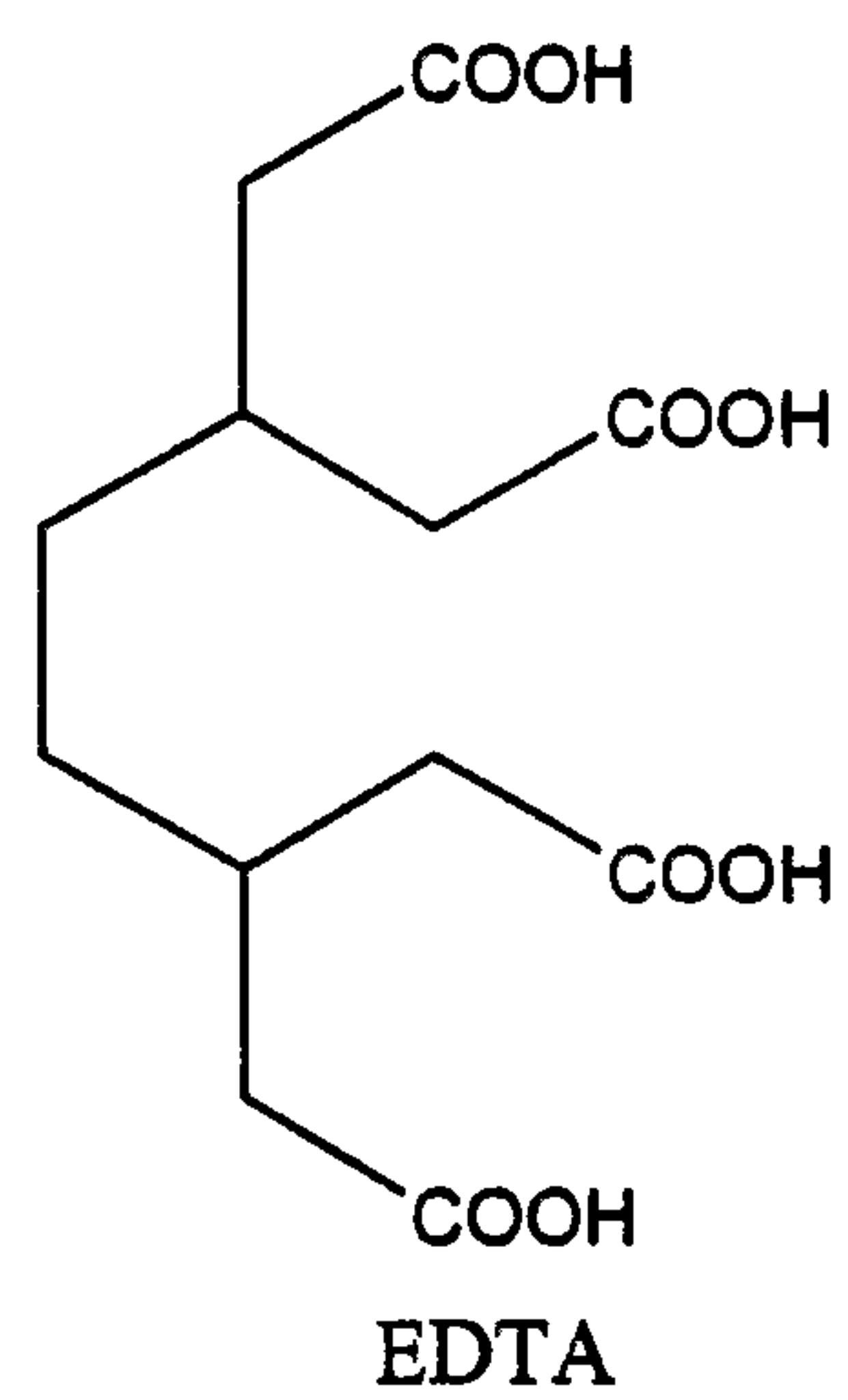
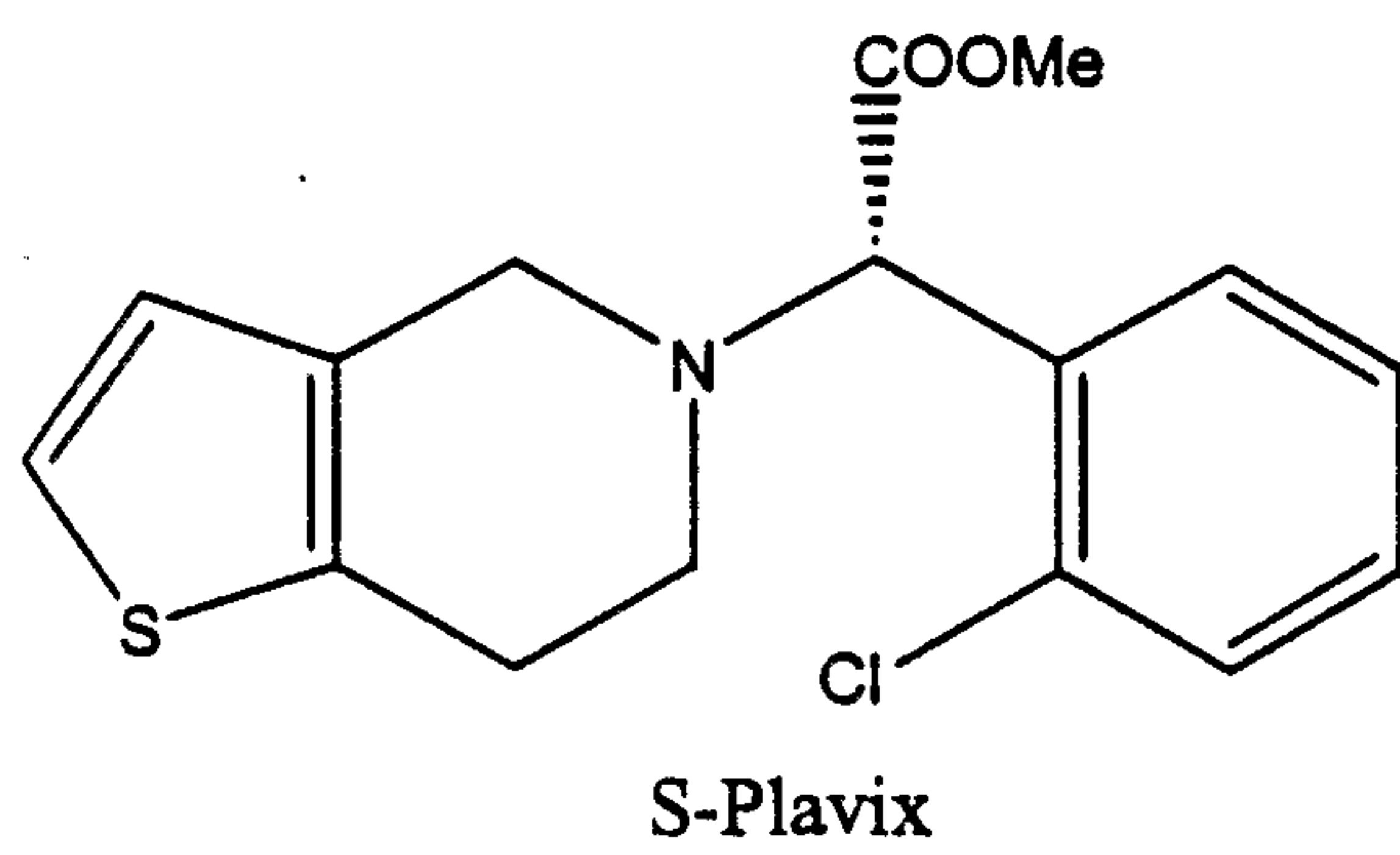
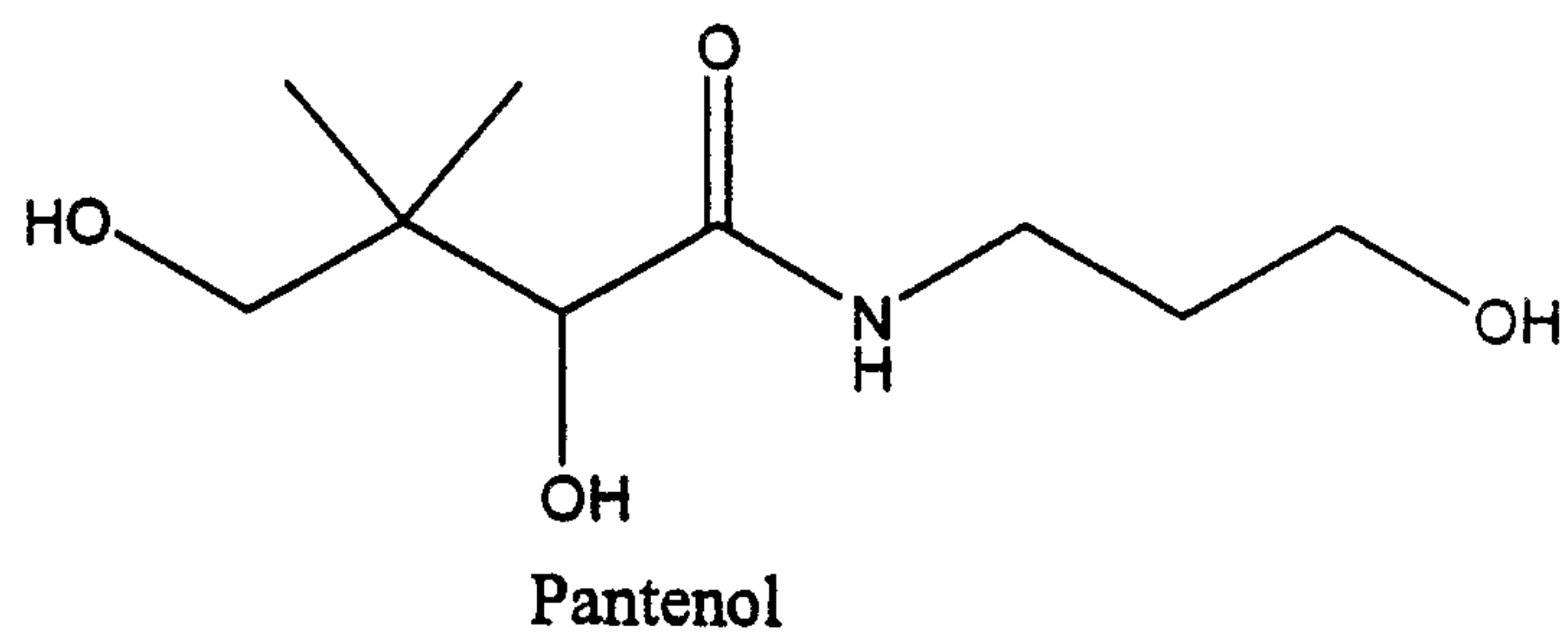
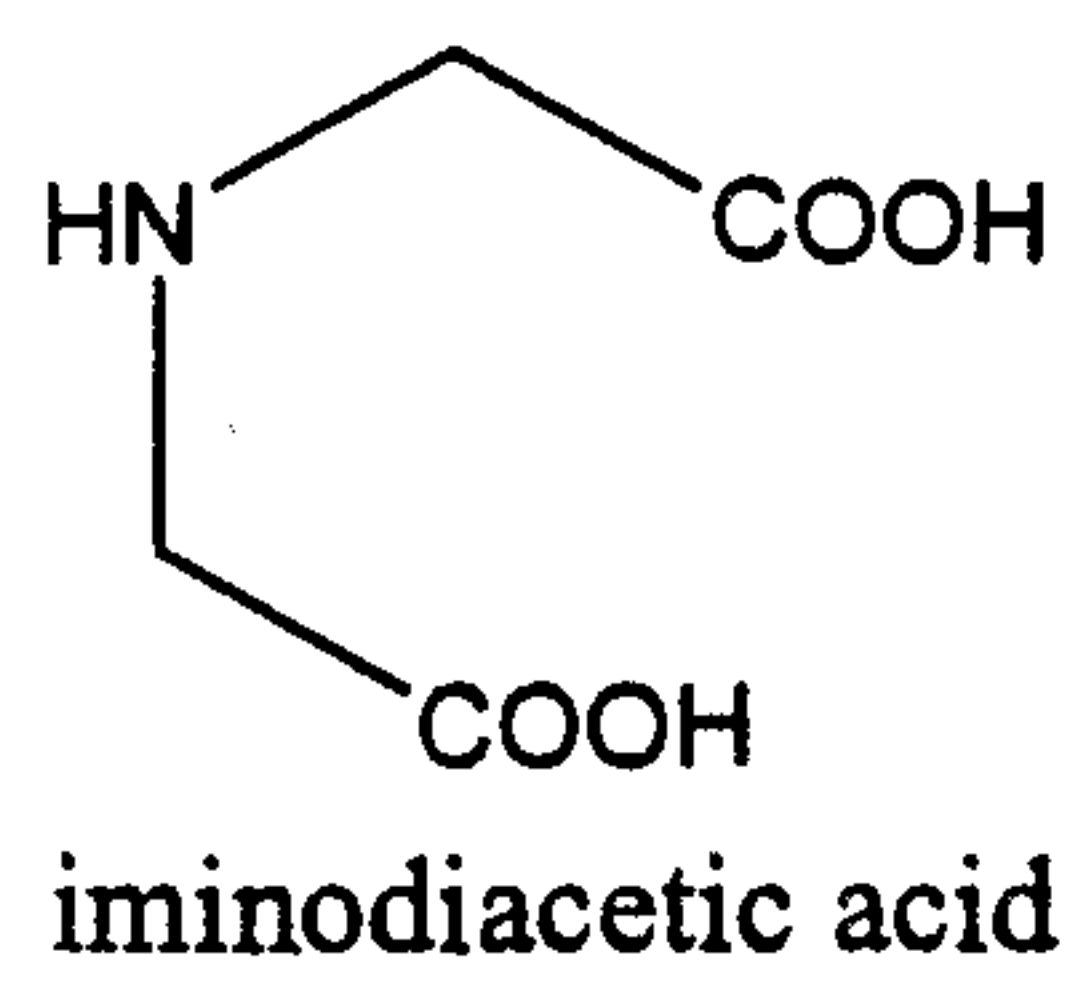




Figure 15

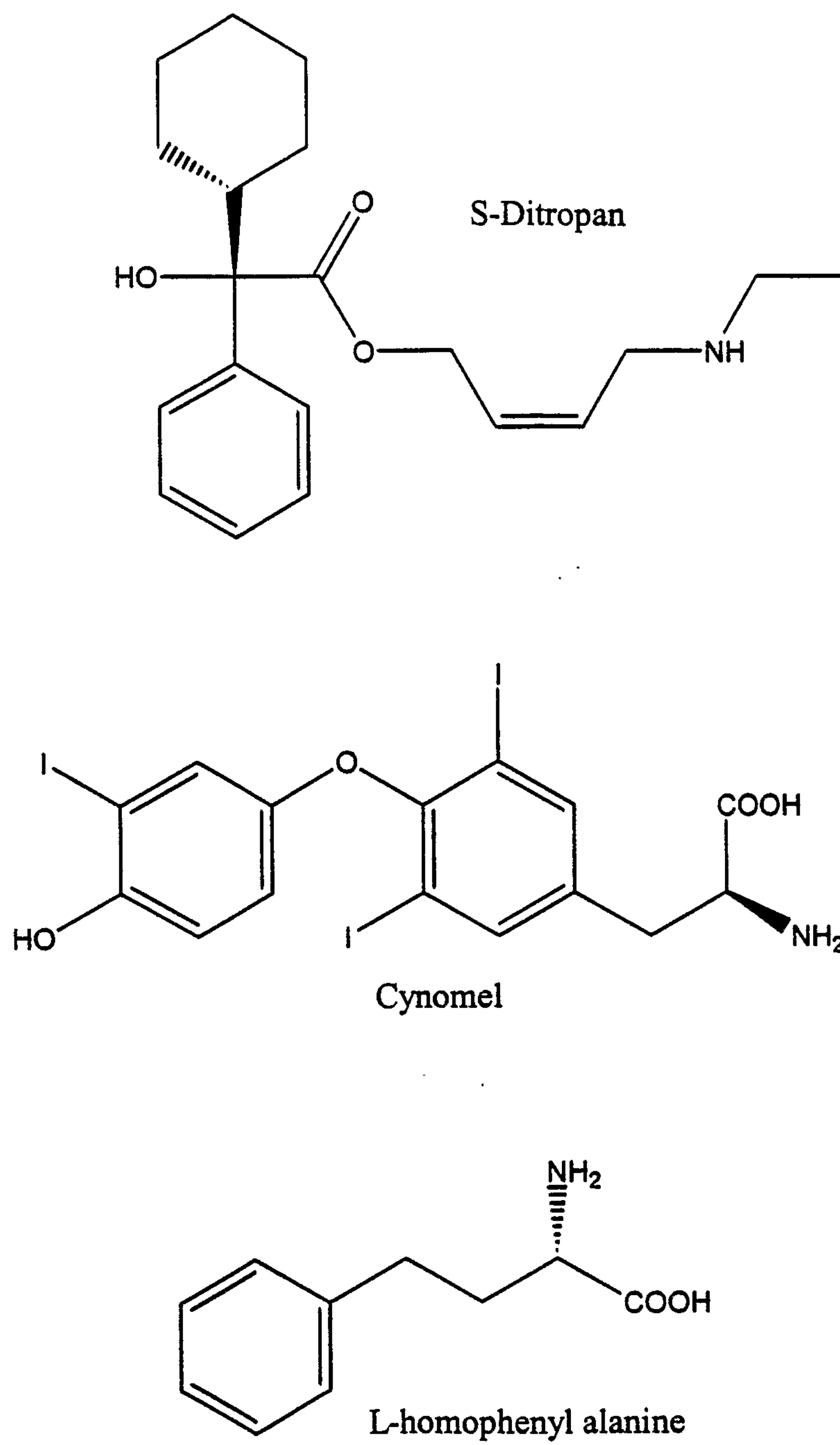
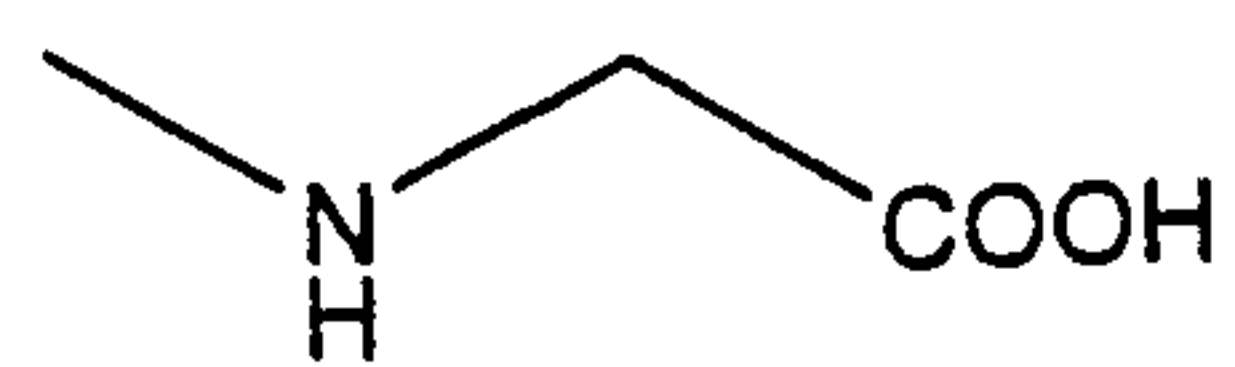
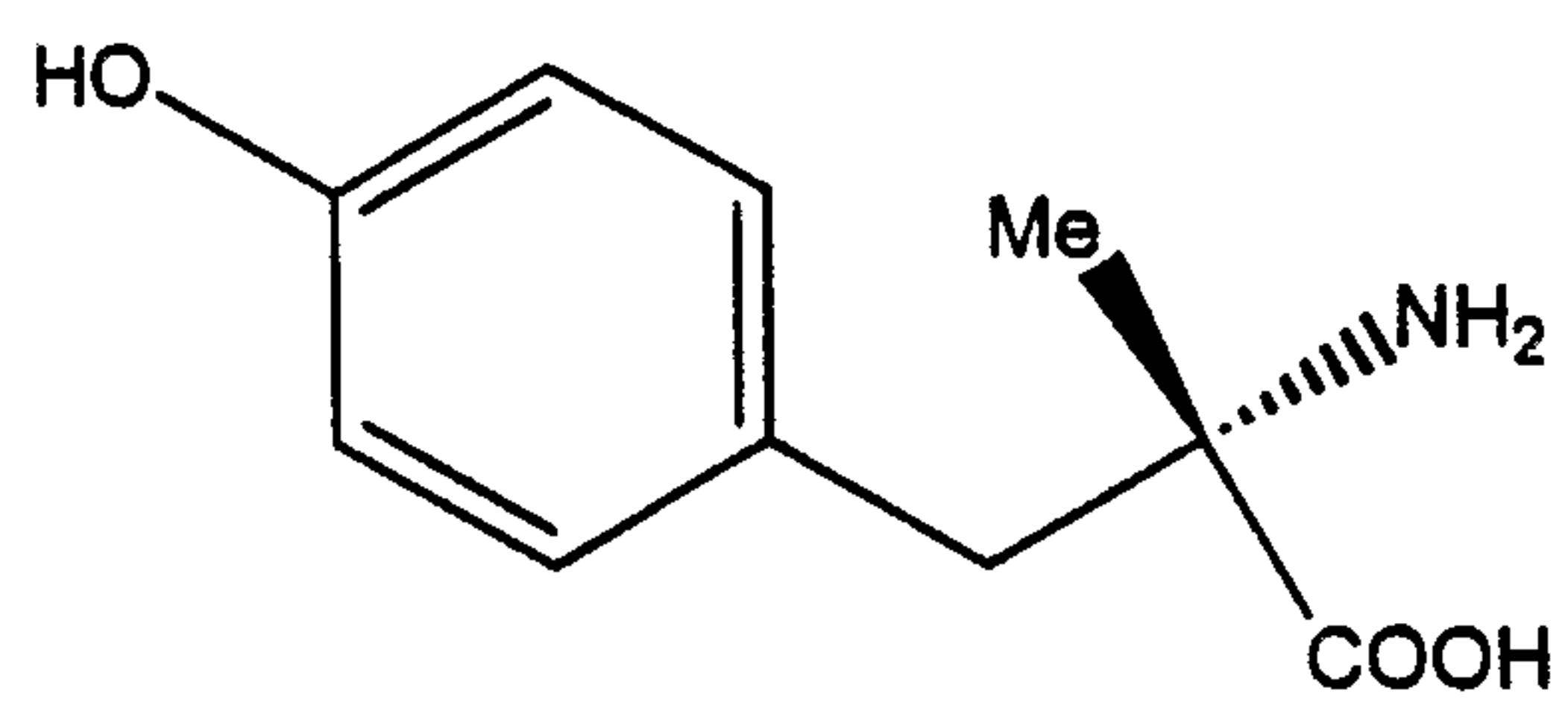




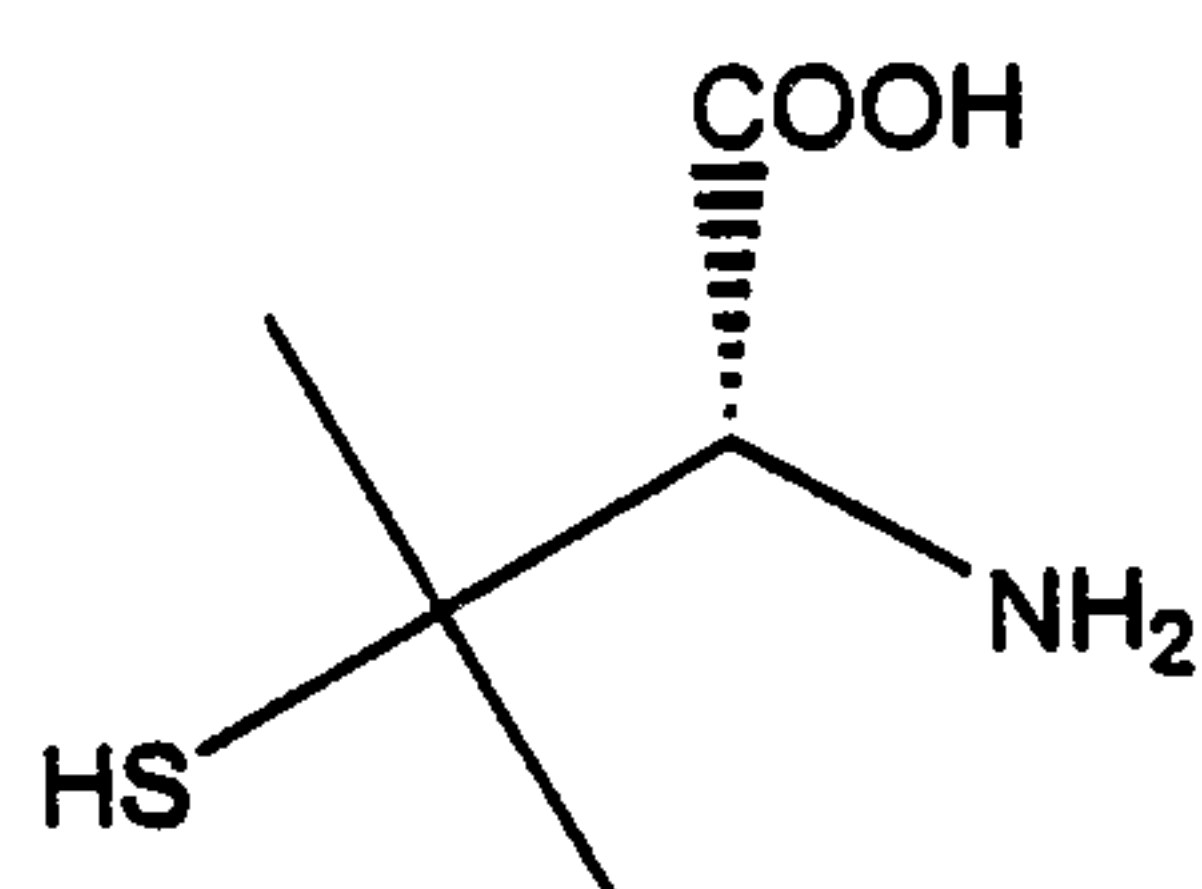
Figure 15



Sarcosine



alpha-methyl-L-tyrosine



Cuprimine/Depen



Figure 15

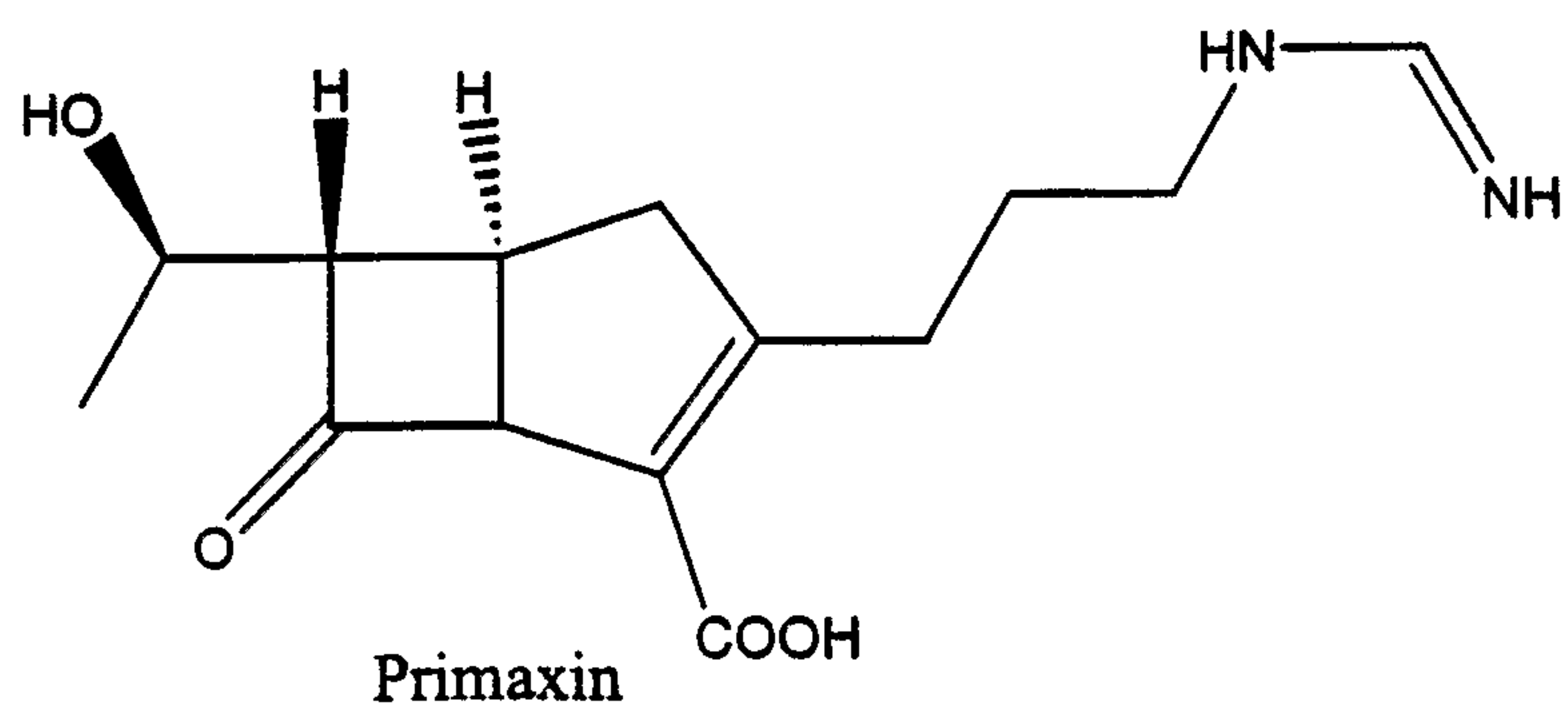
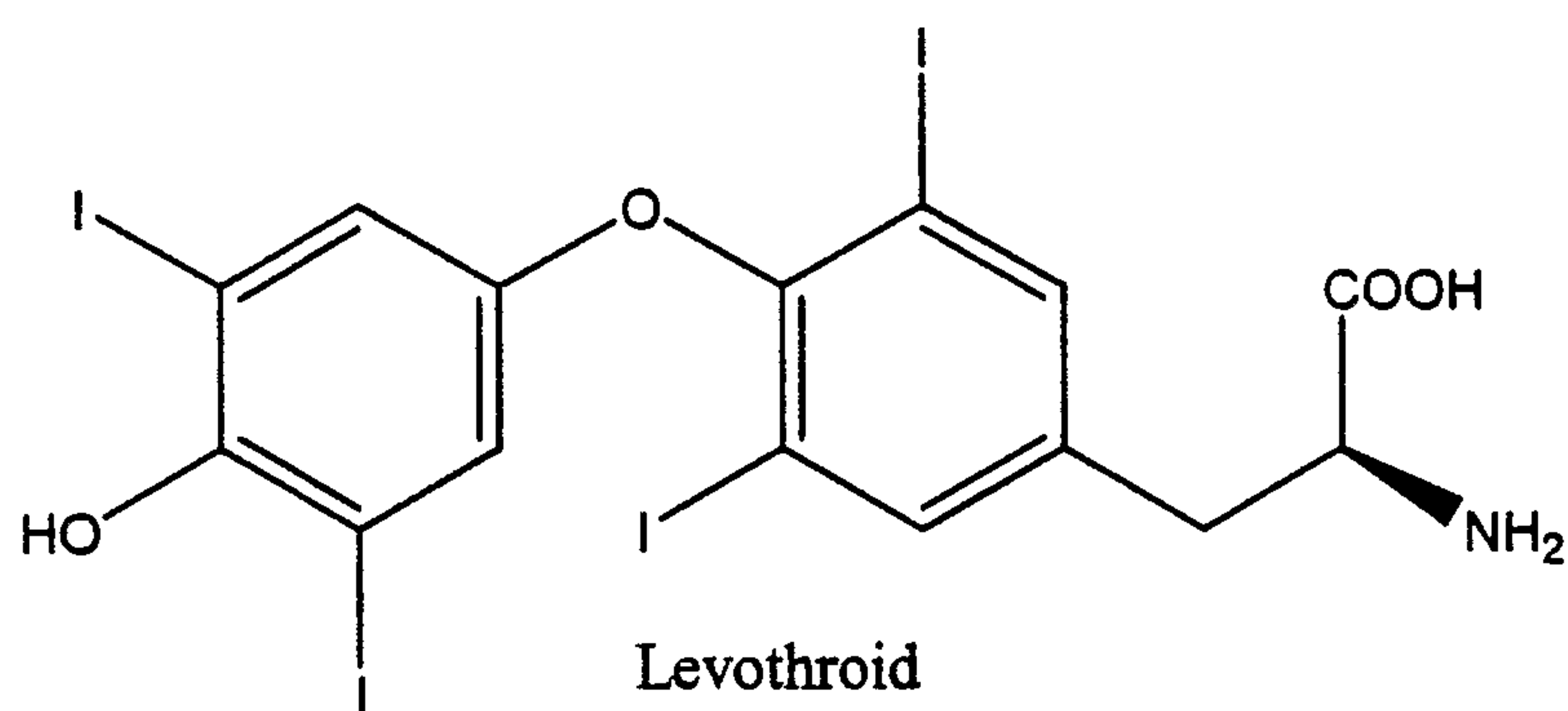
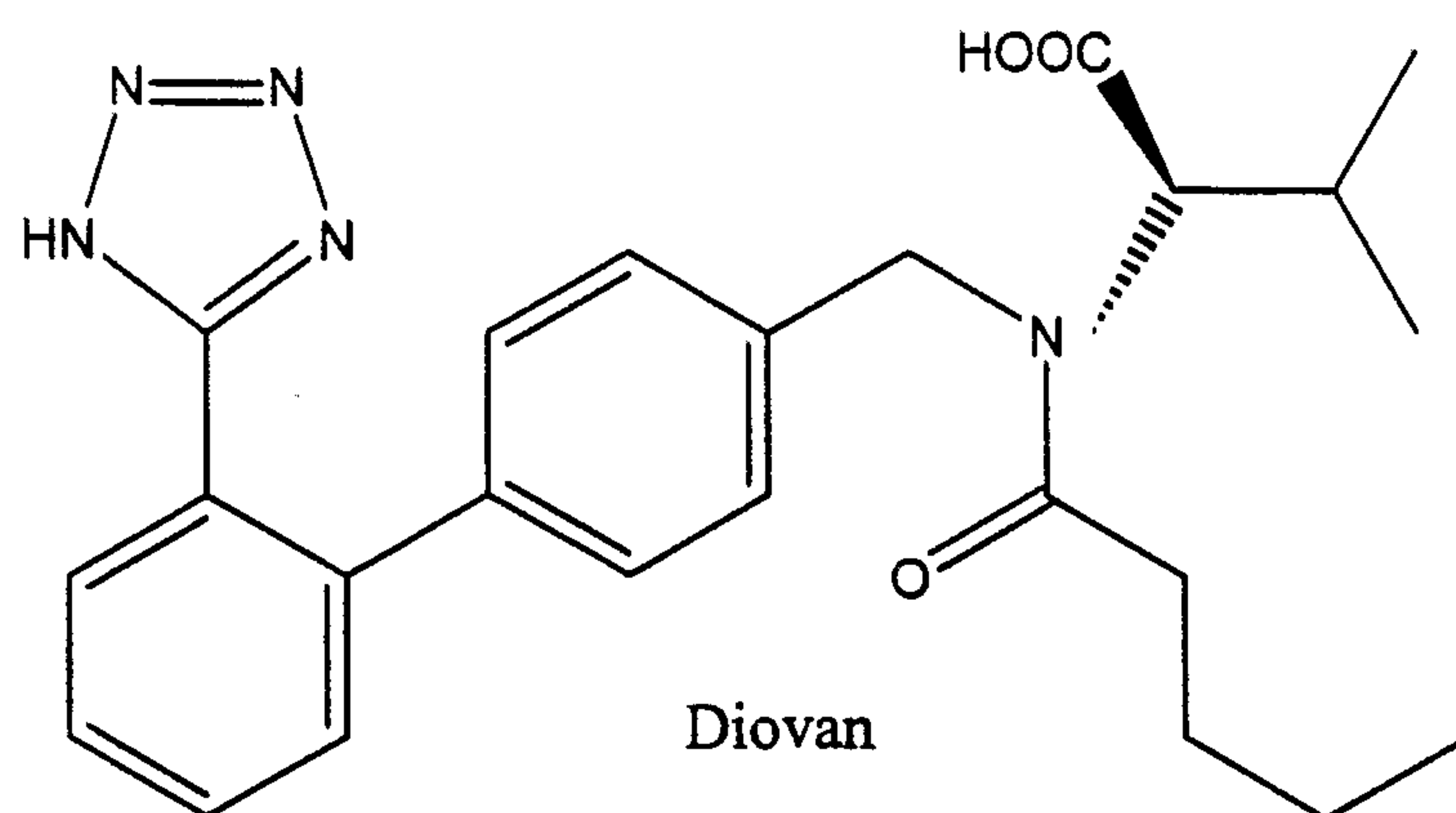
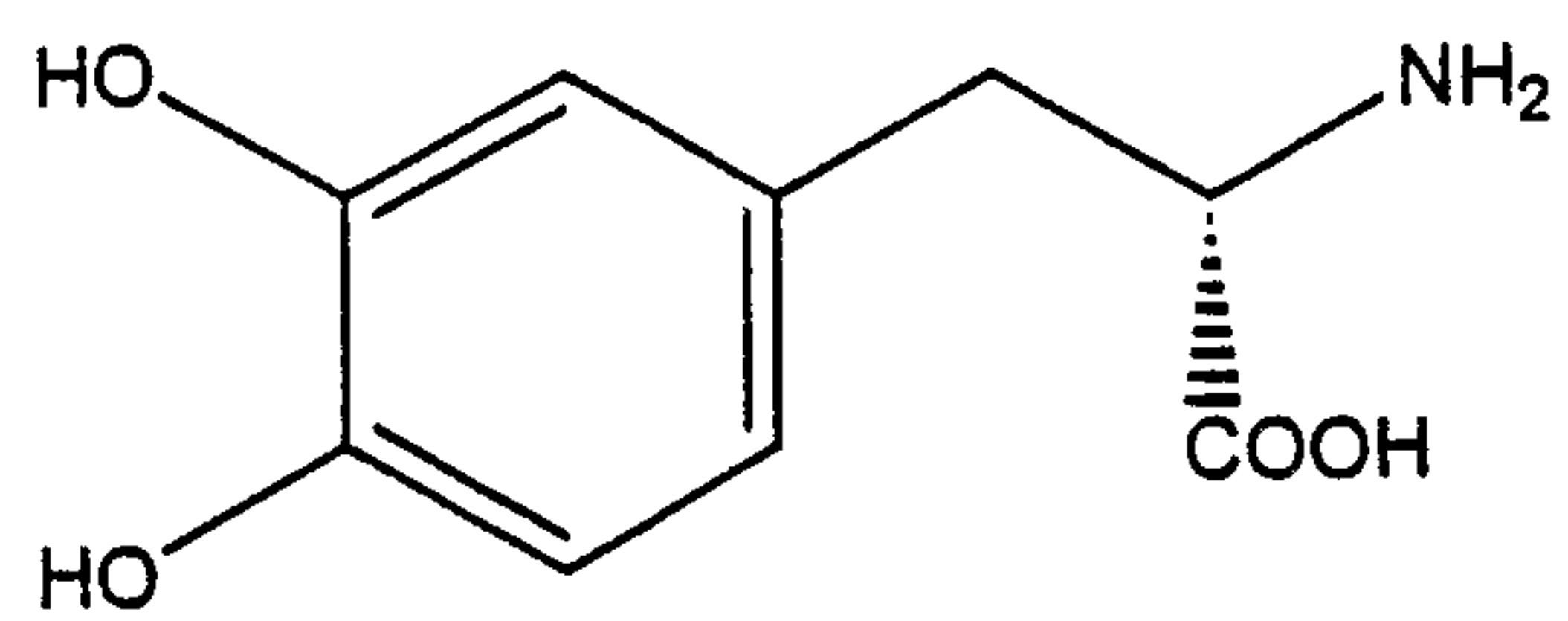
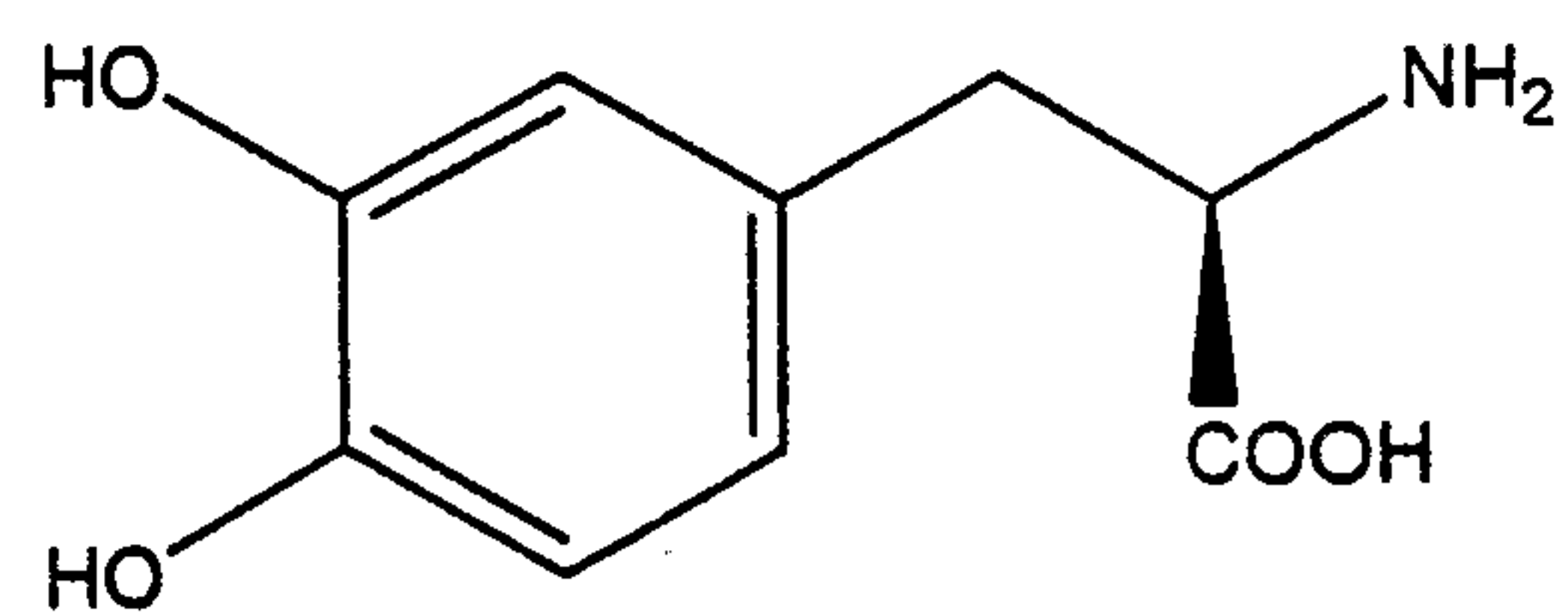




Figure 15



L-Dopa



D-Dopa

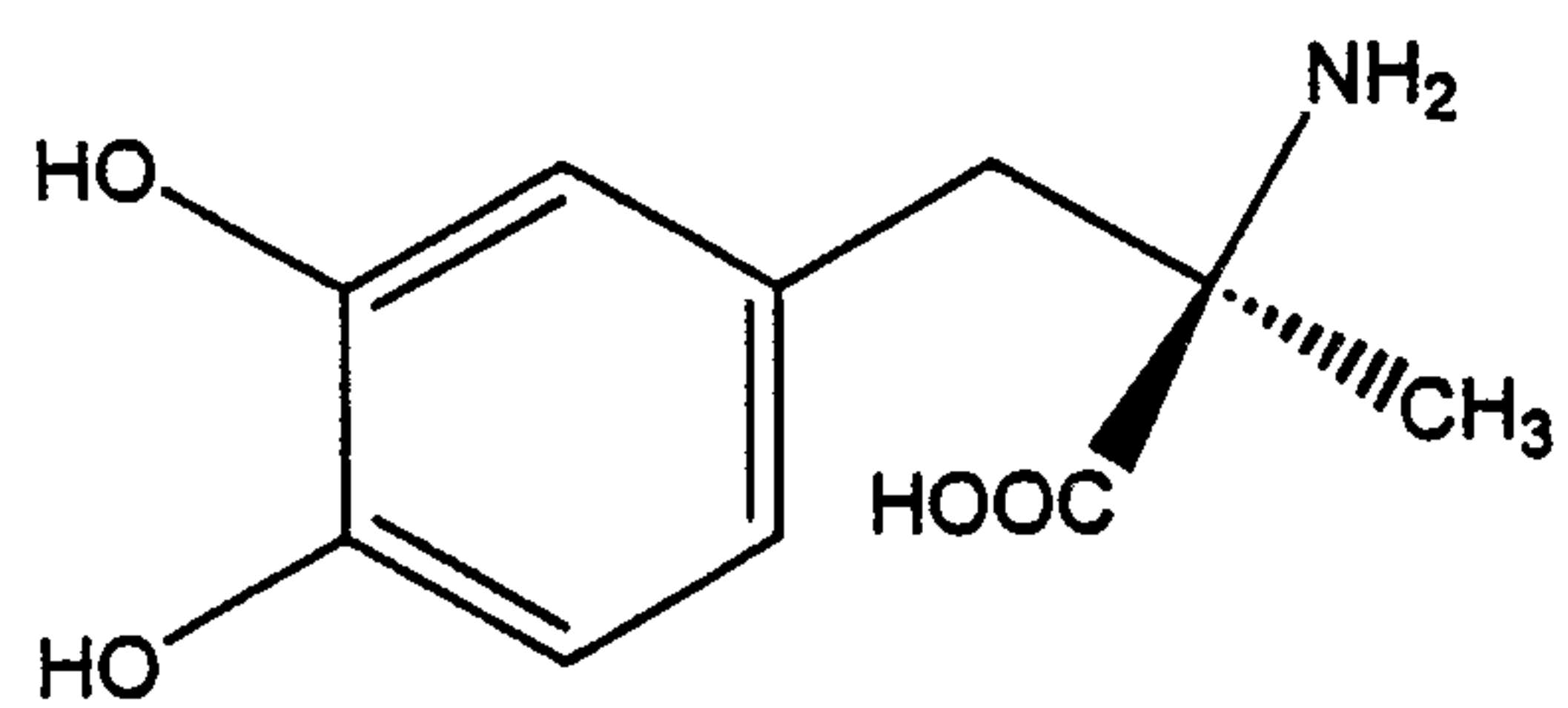
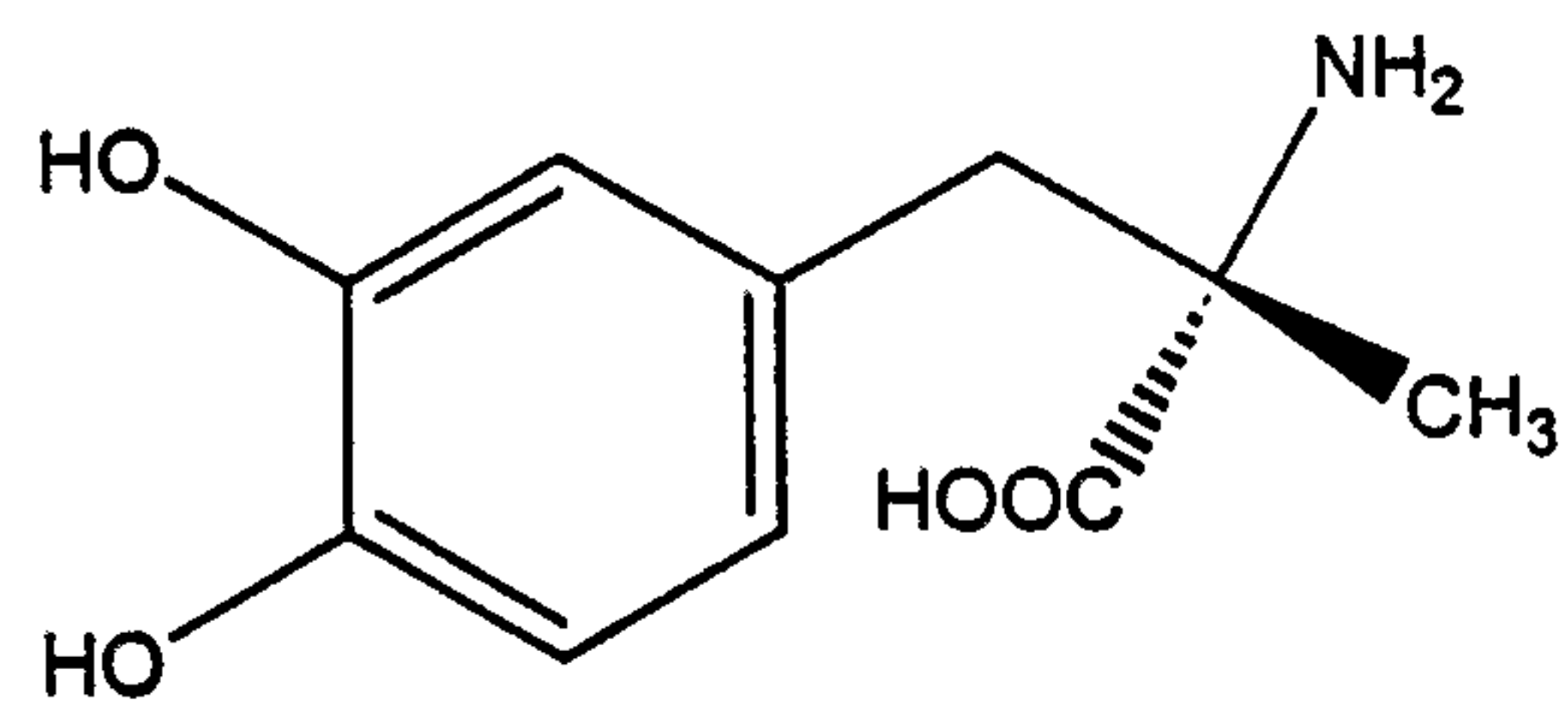
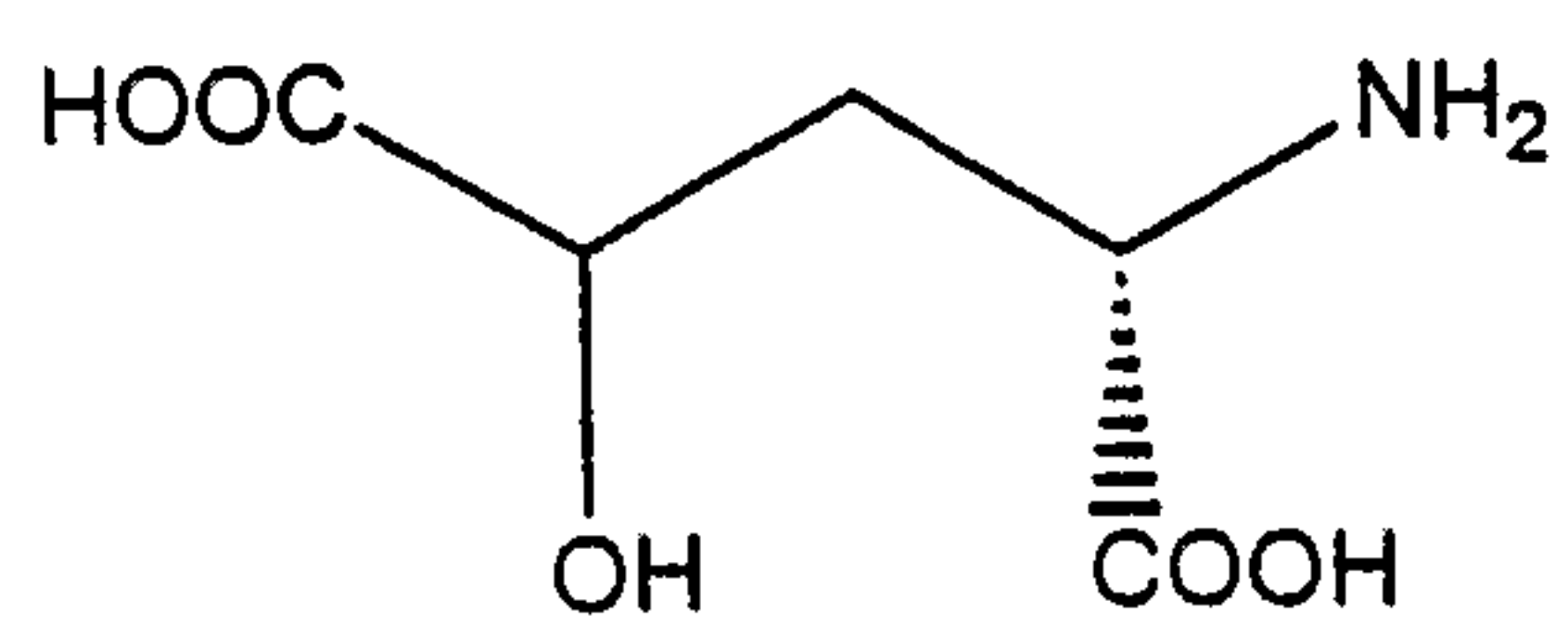
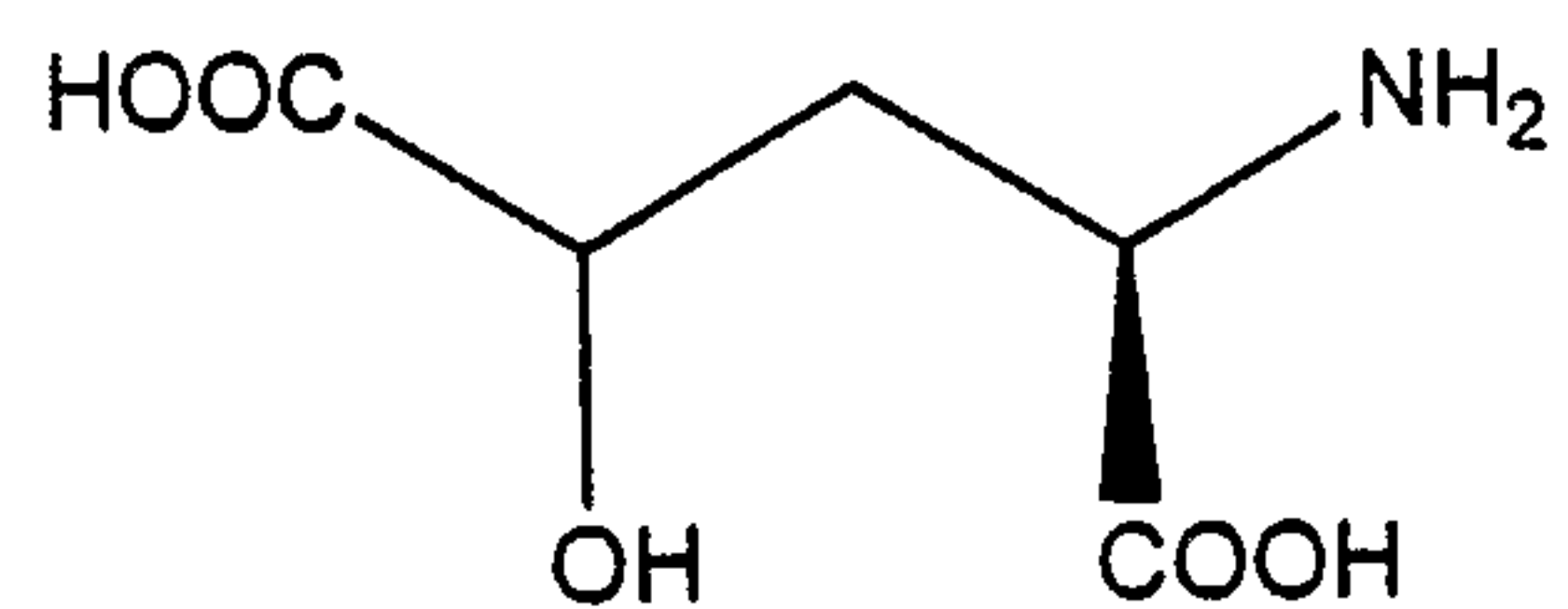
D- $\alpha$ -methyl-DOPAL- $\alpha$ -methyl-DOPA



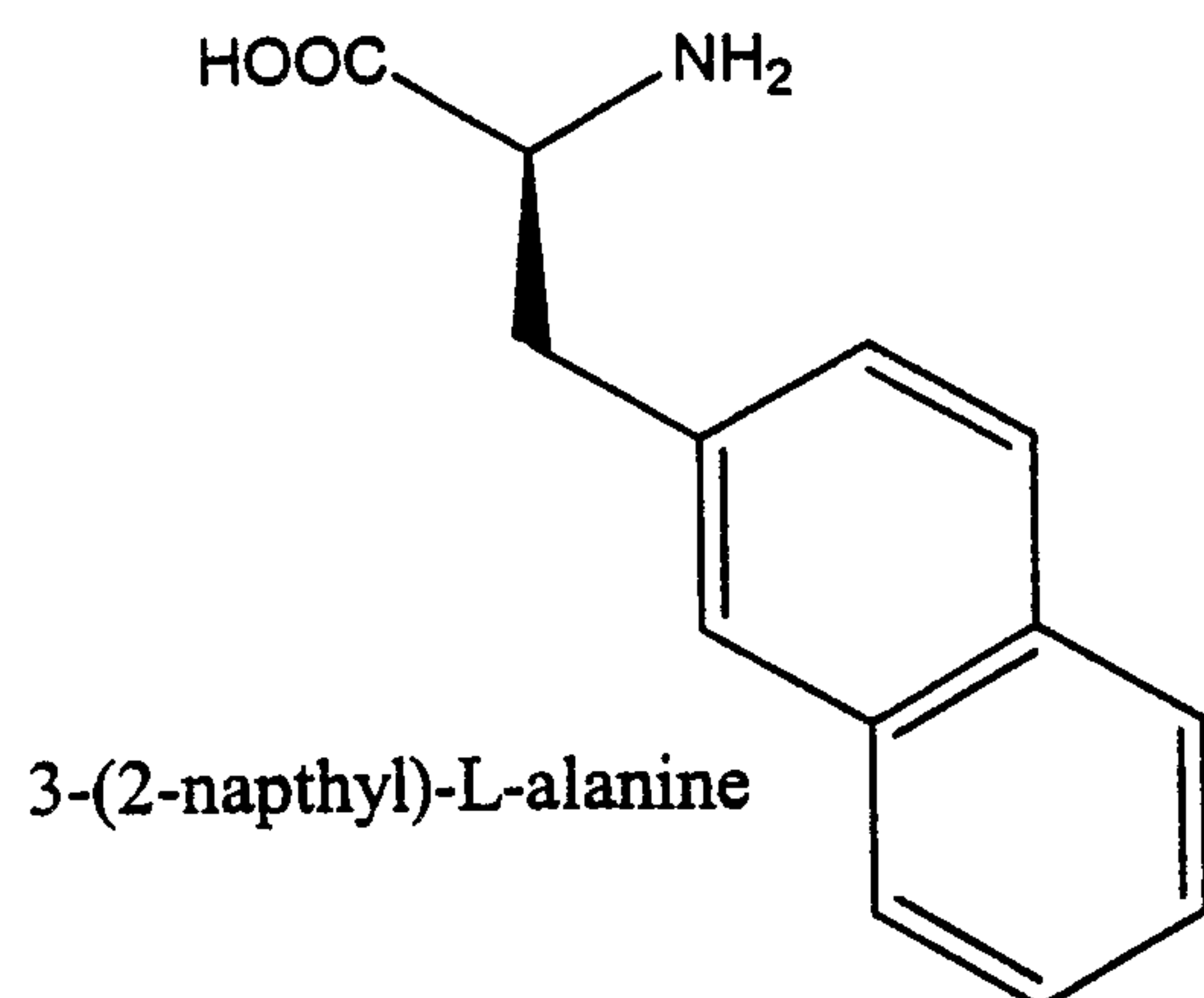
Figure 15



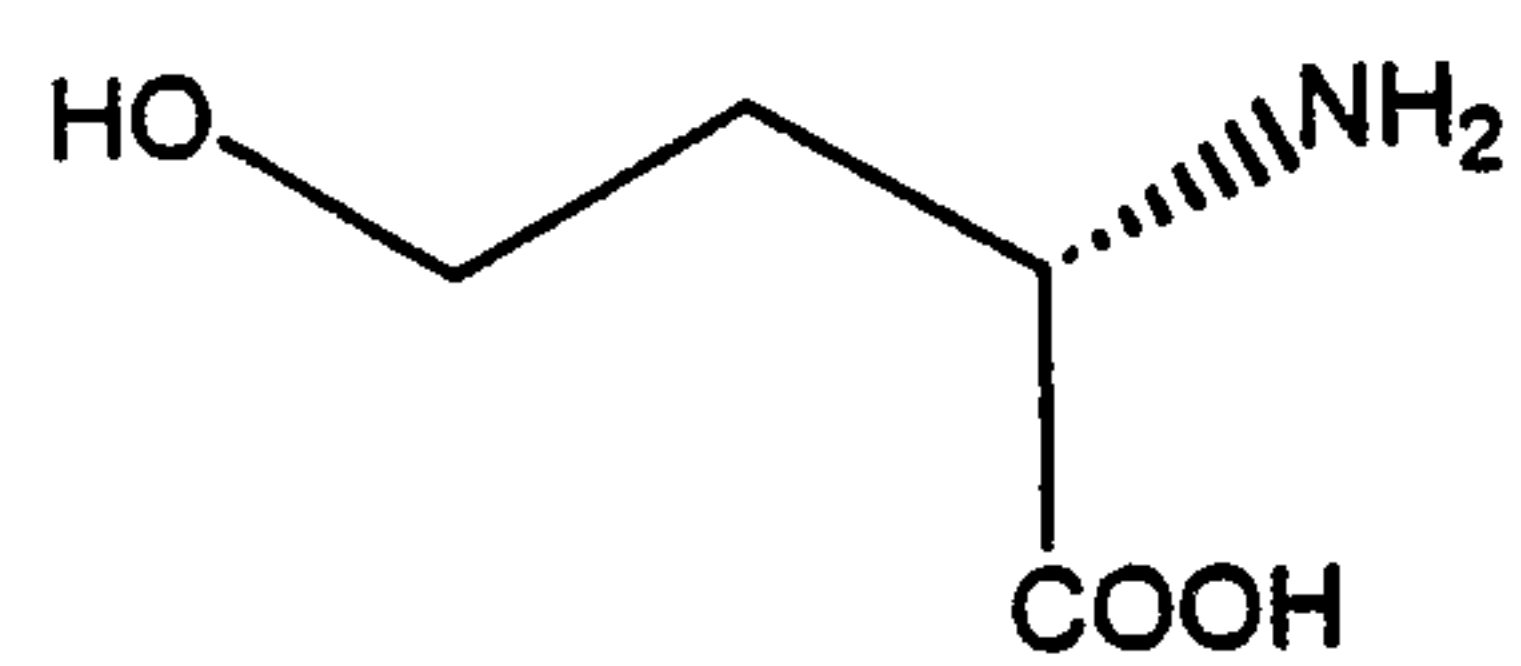
L-gamma-hydroxy glutamate



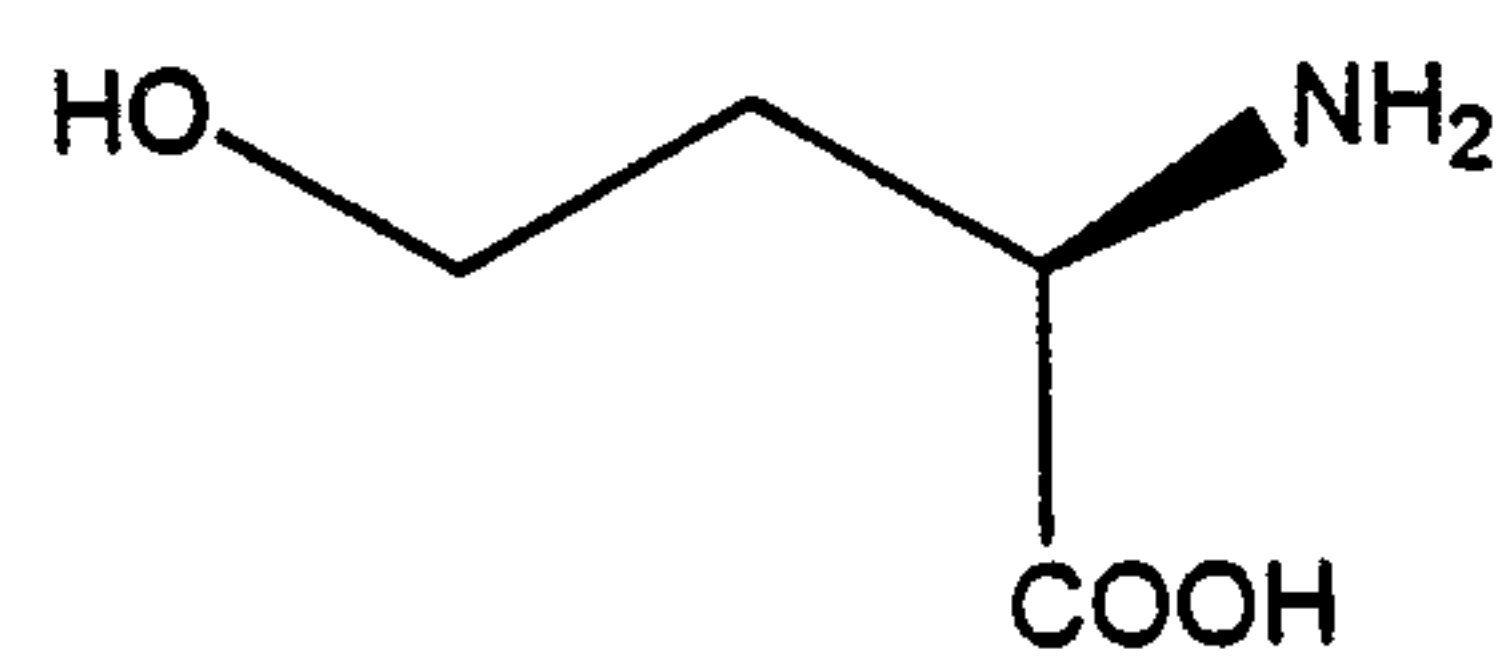
D-gamma-hydroxy glutamate



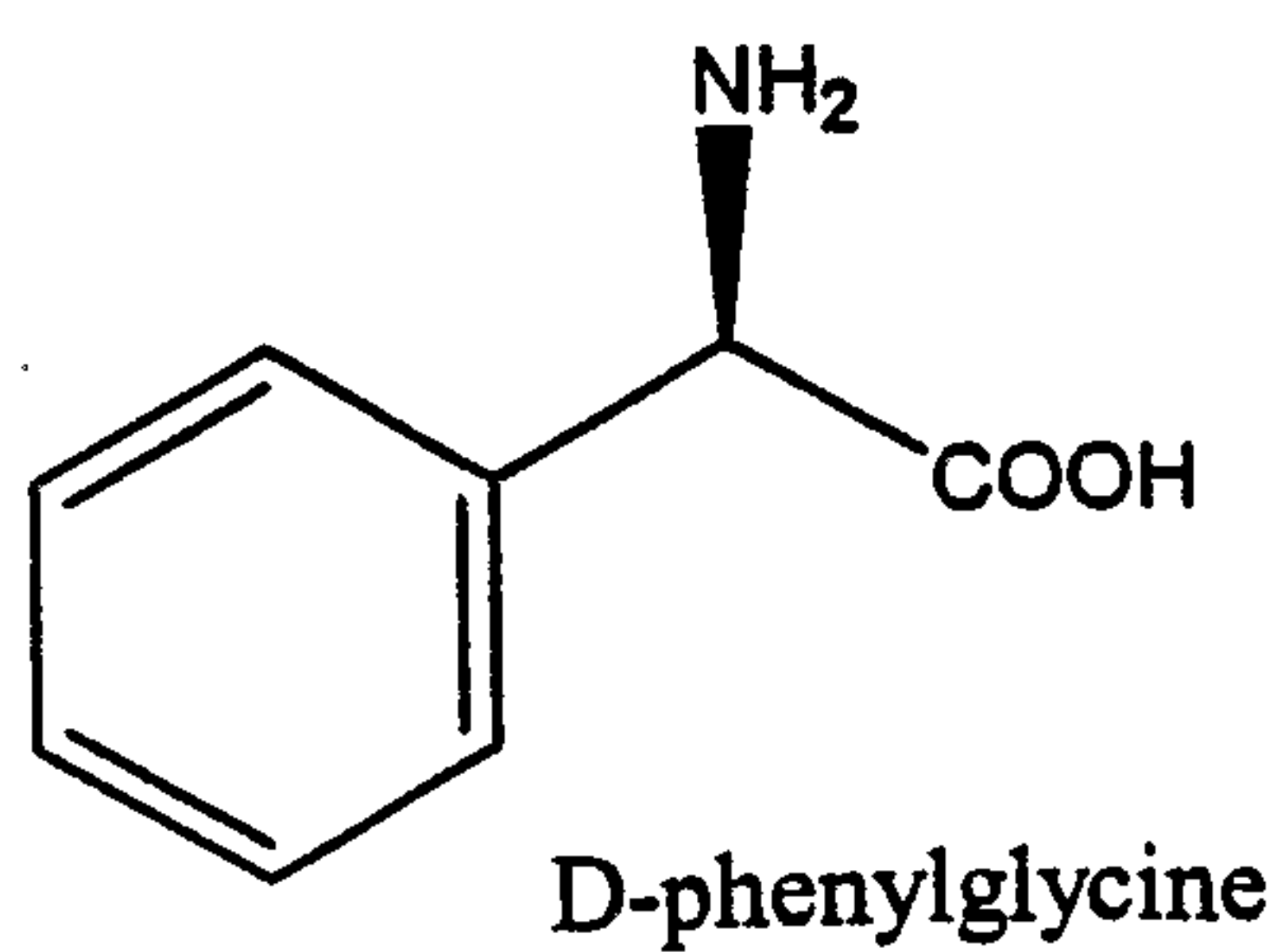
3-(2-naphthyl)-L-alanine



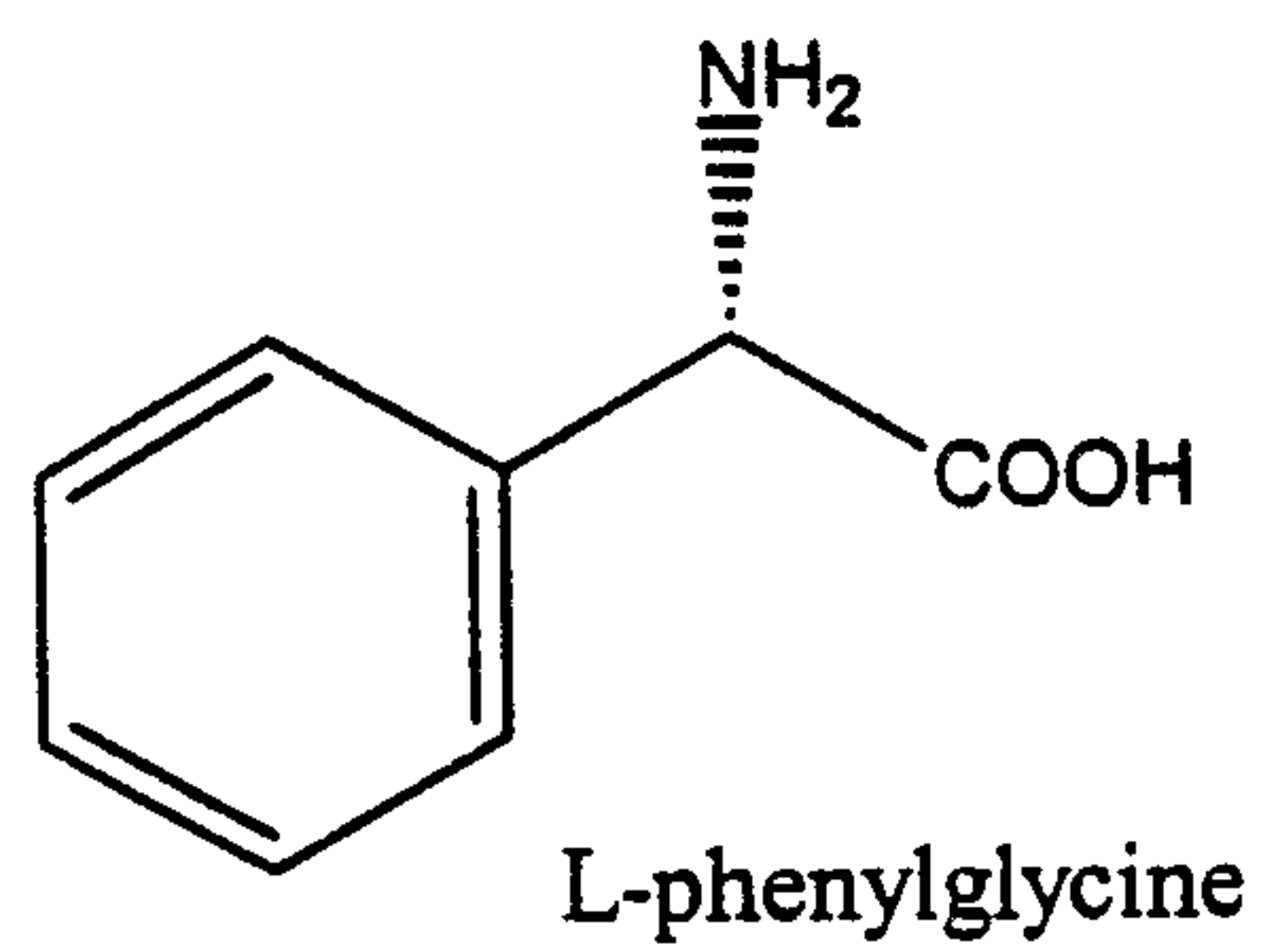
L-homoserine



D-homoserine



D-phenylglycine



L-phenylglycine



Figure 15

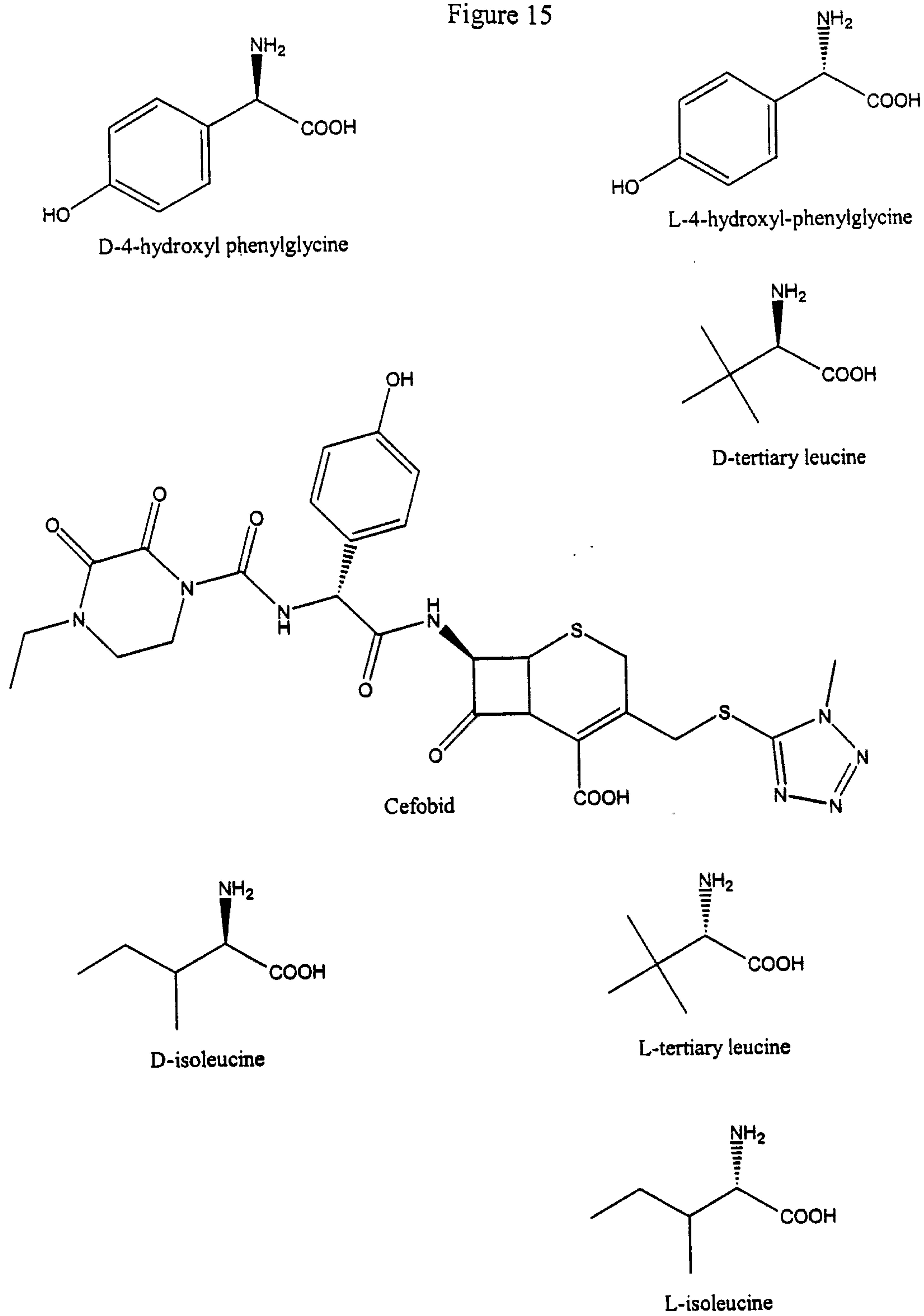




Figure 15

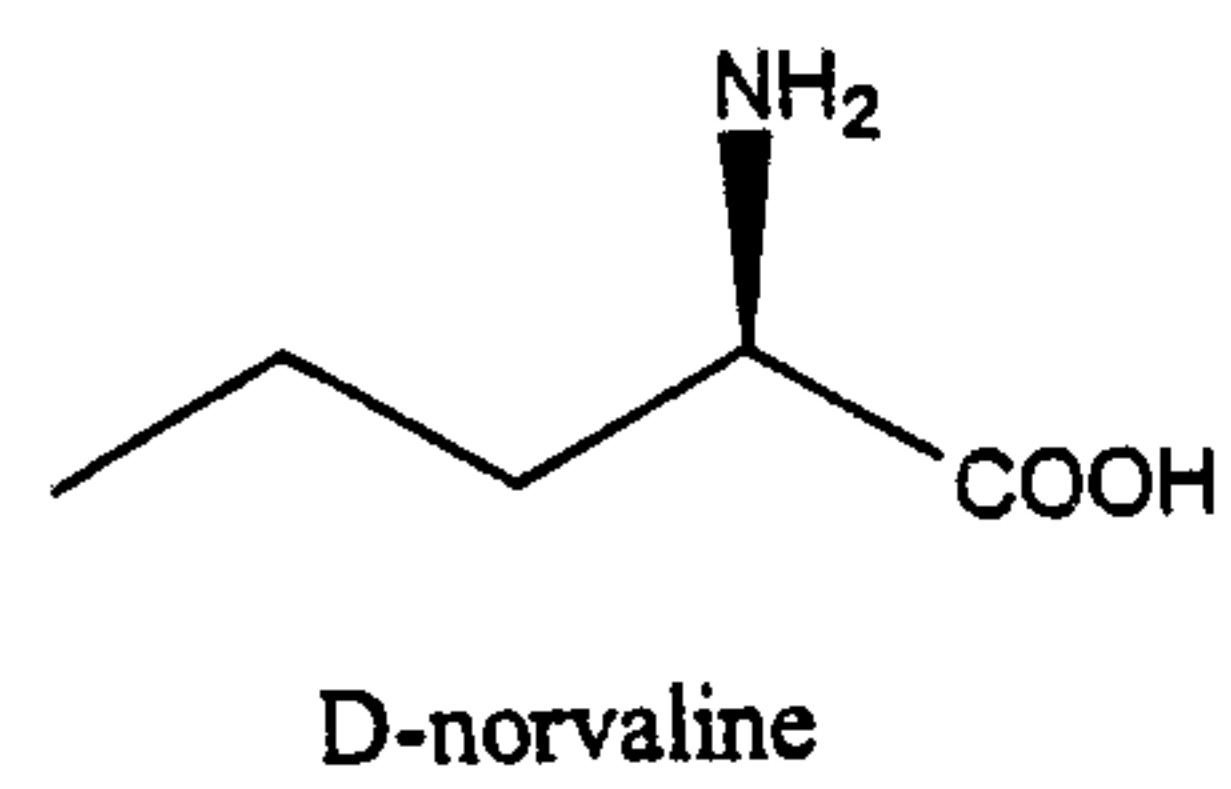
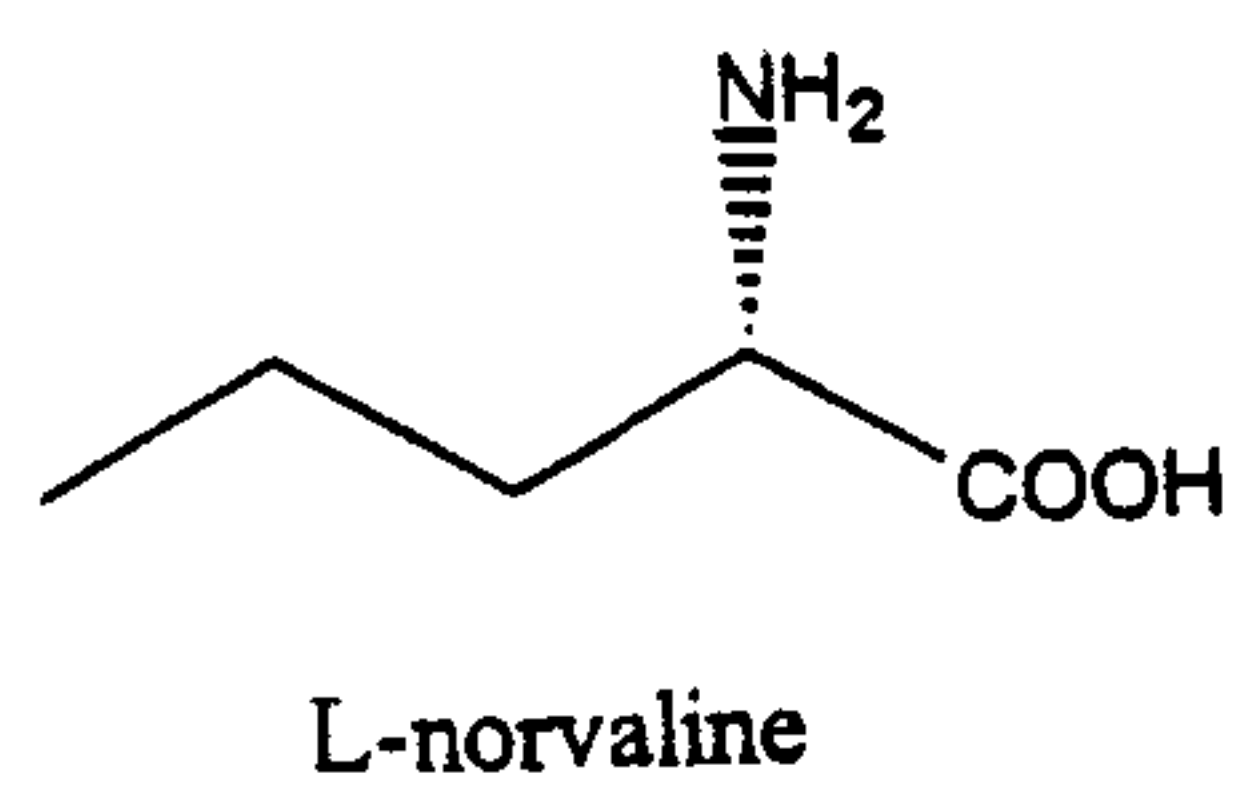
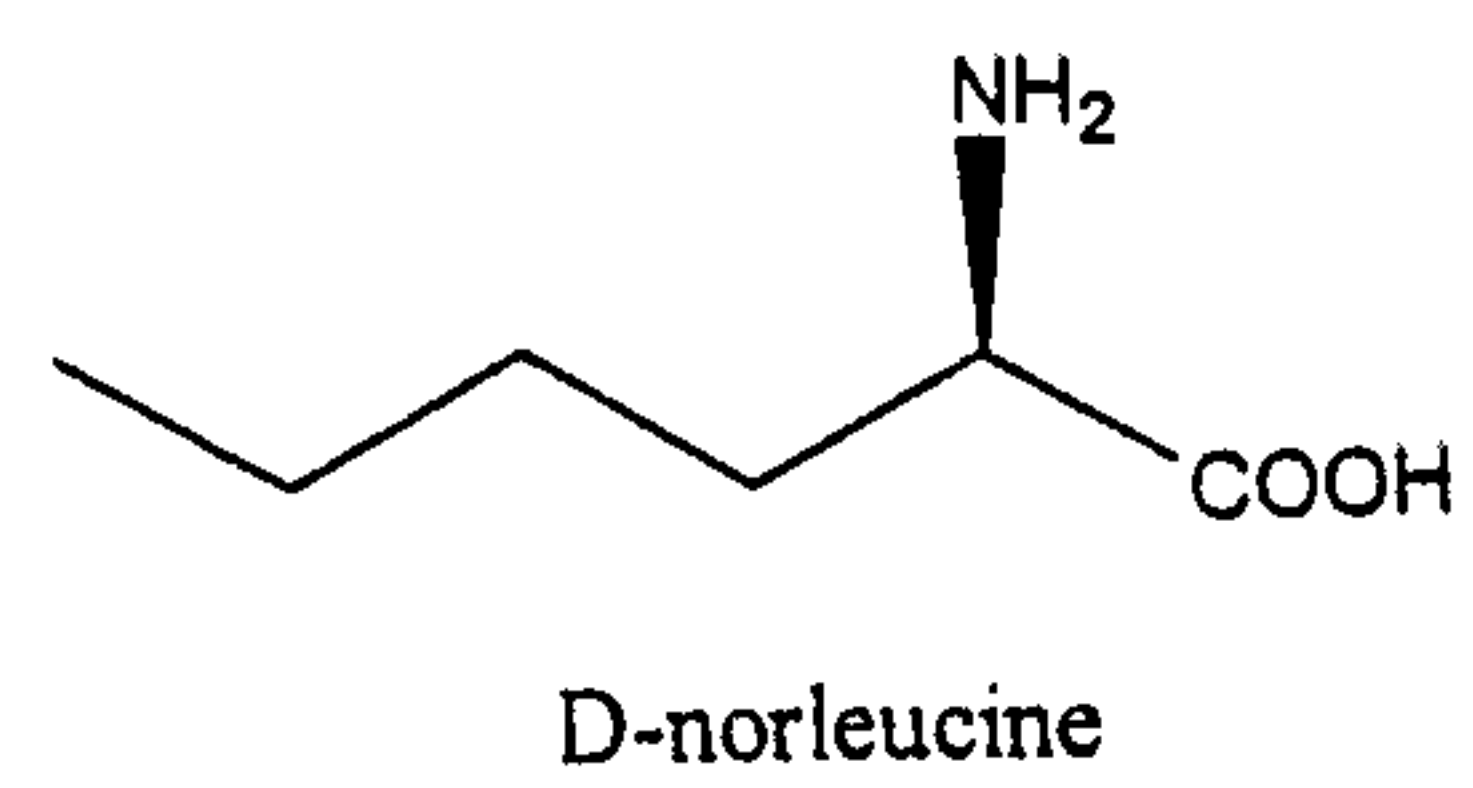
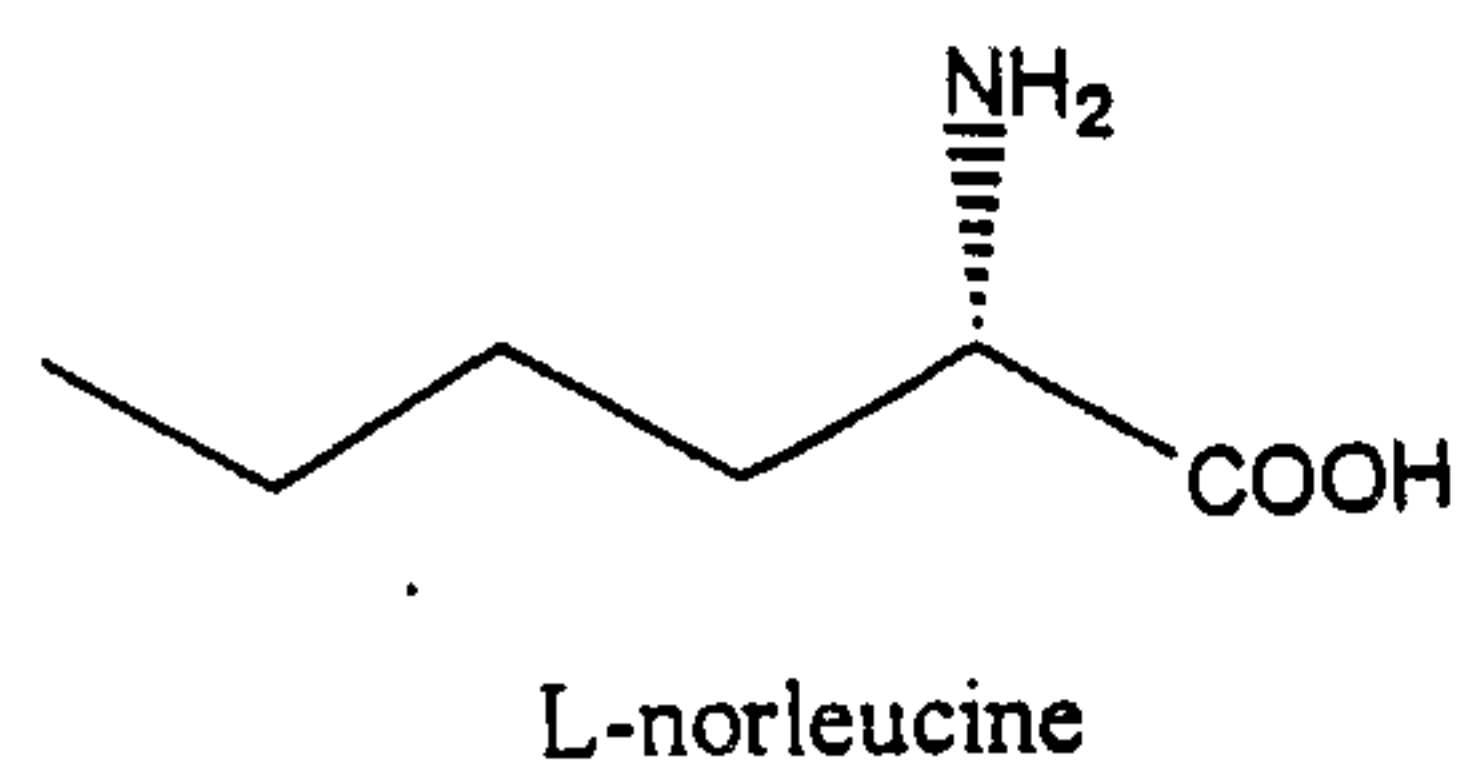




Figure 15

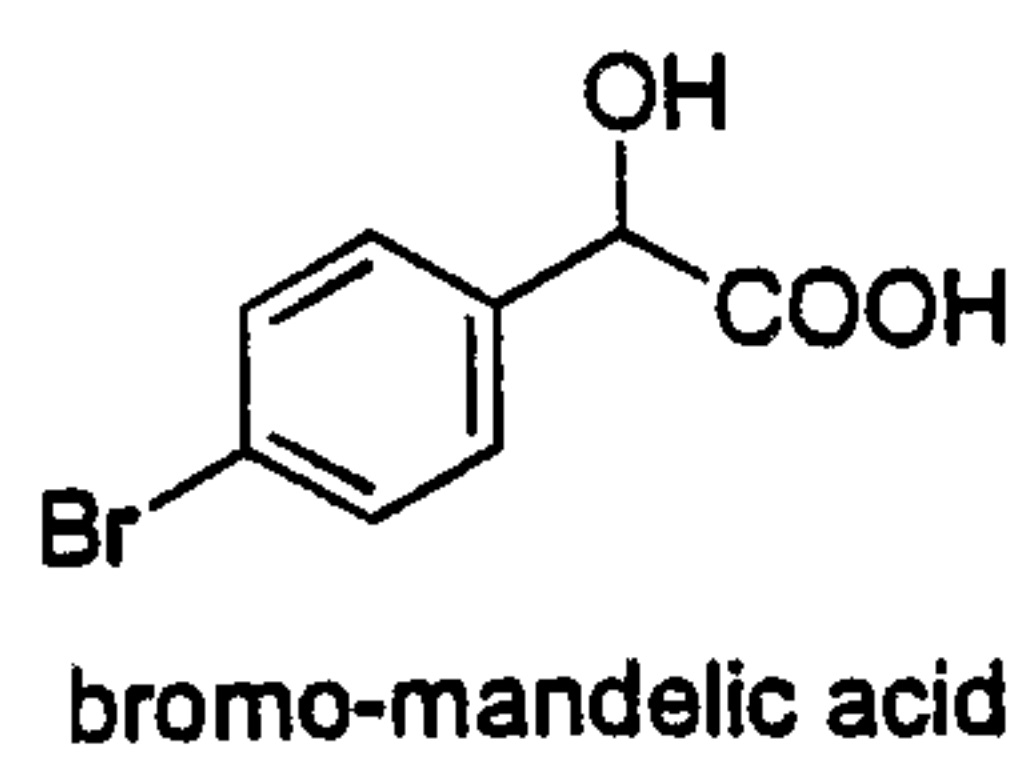
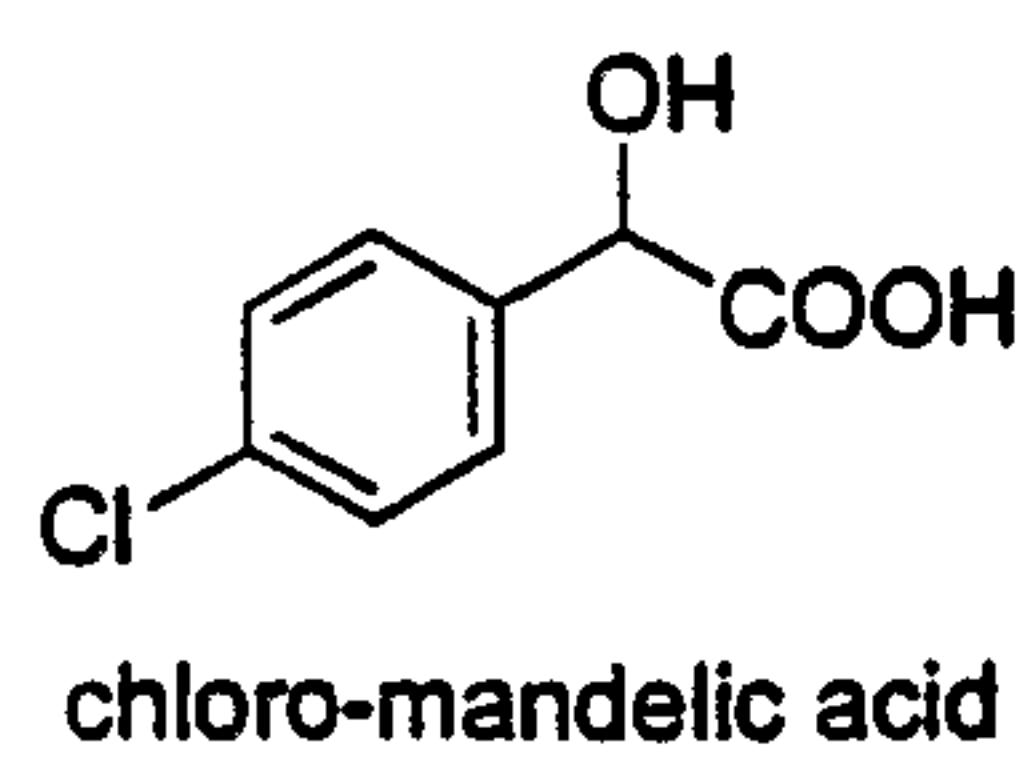
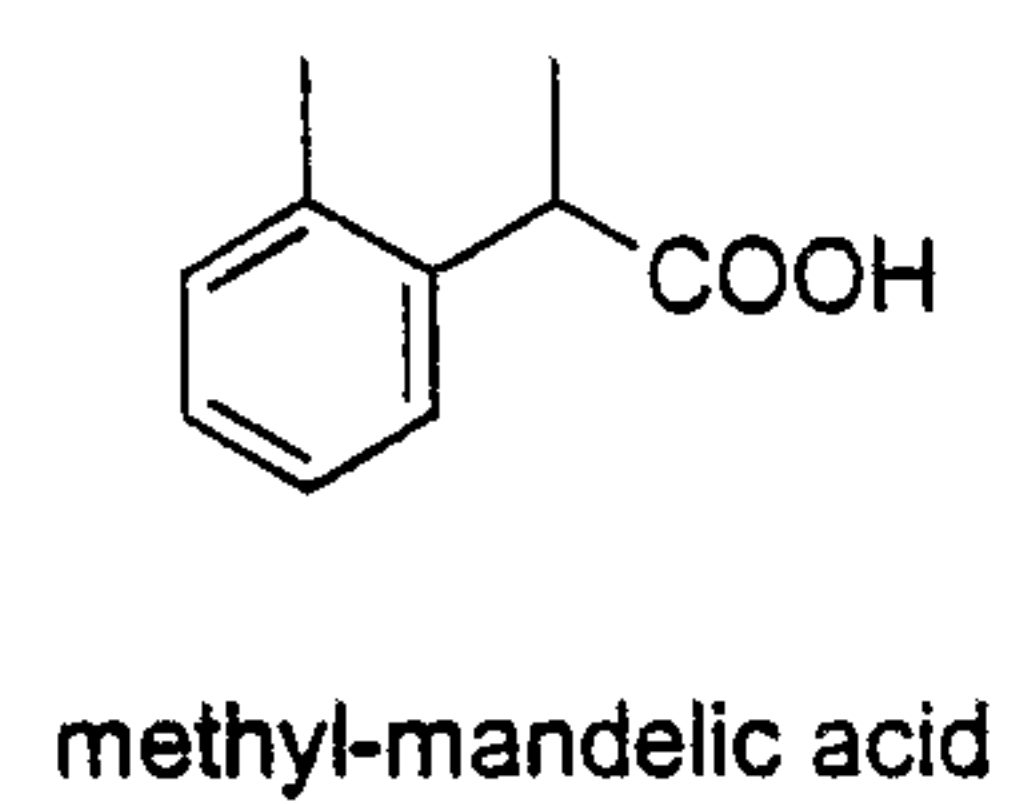
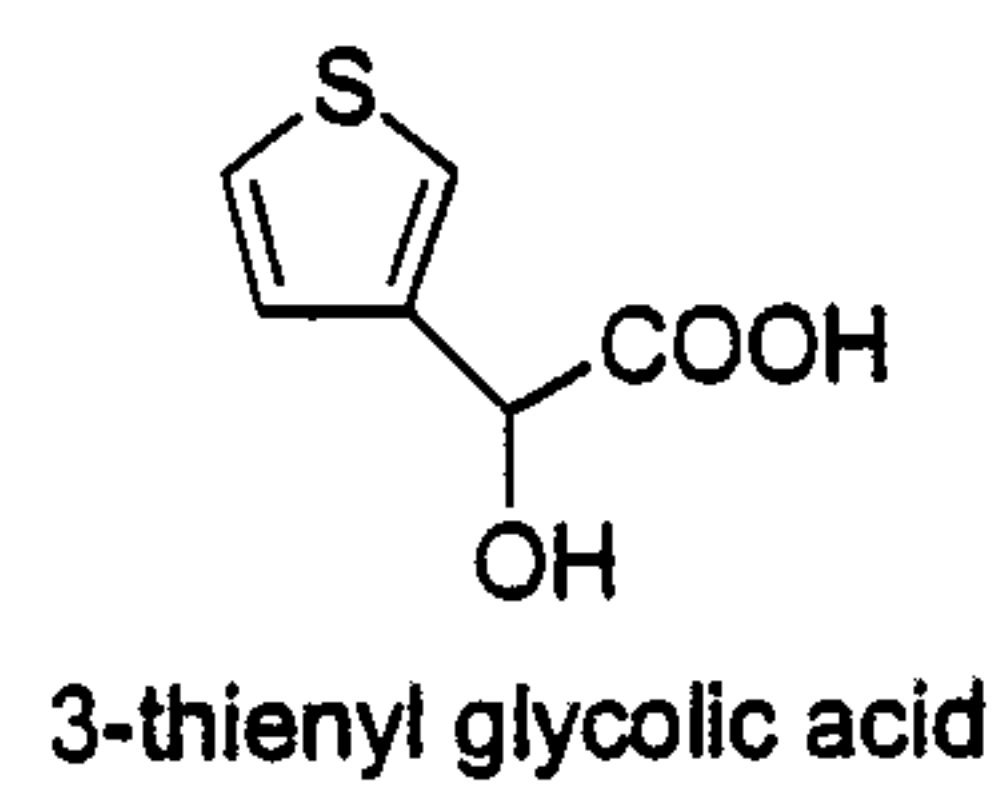
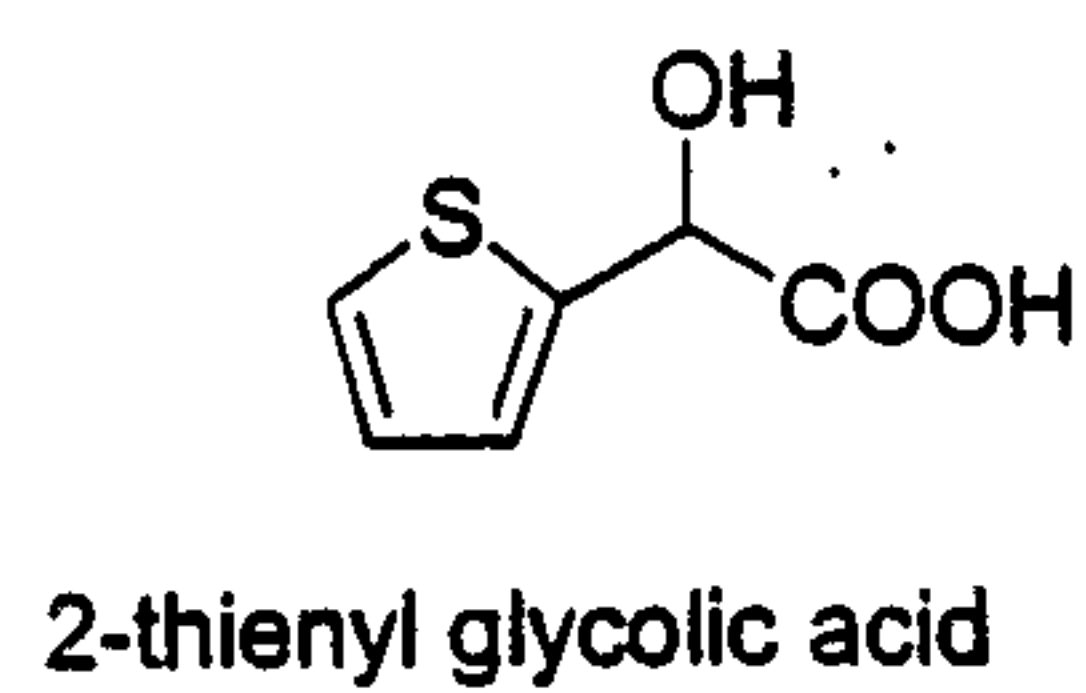
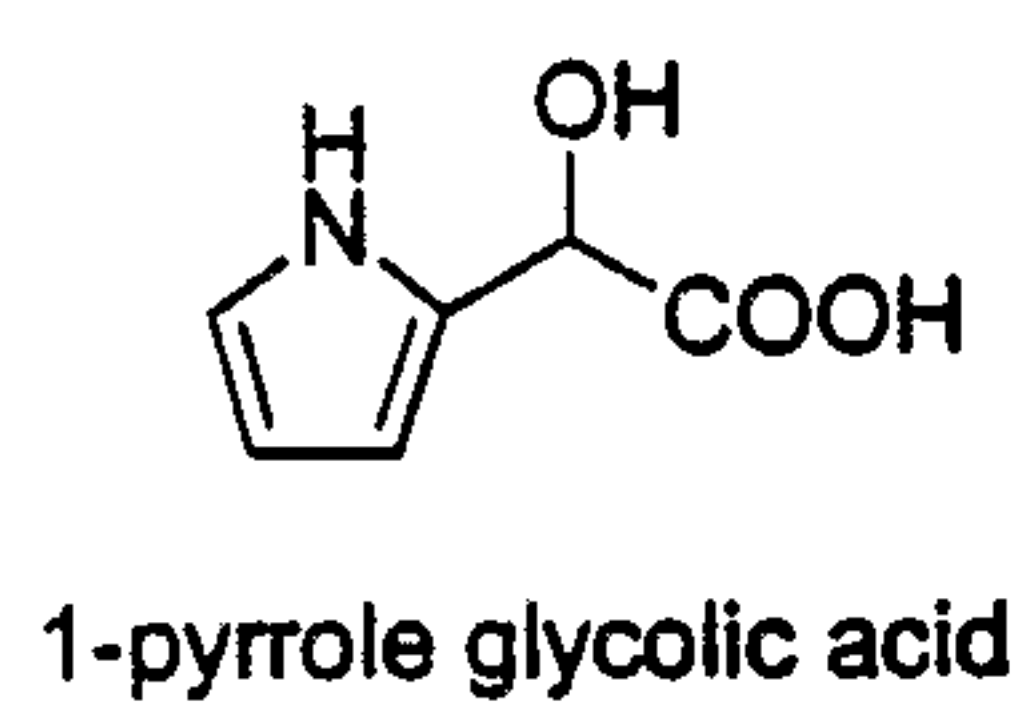
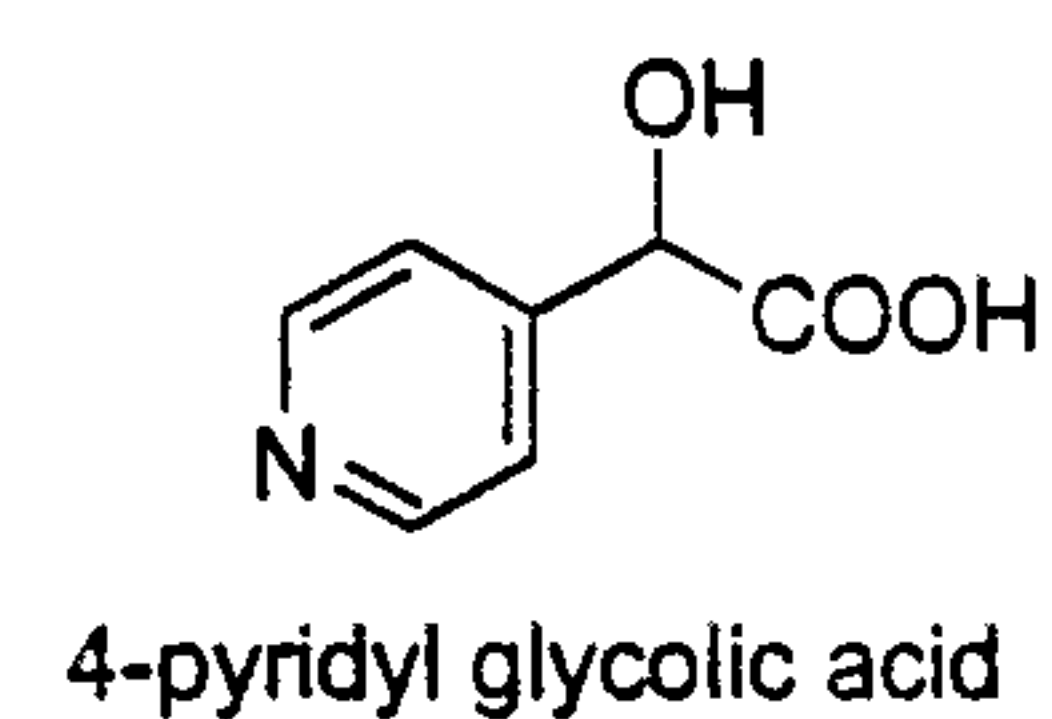
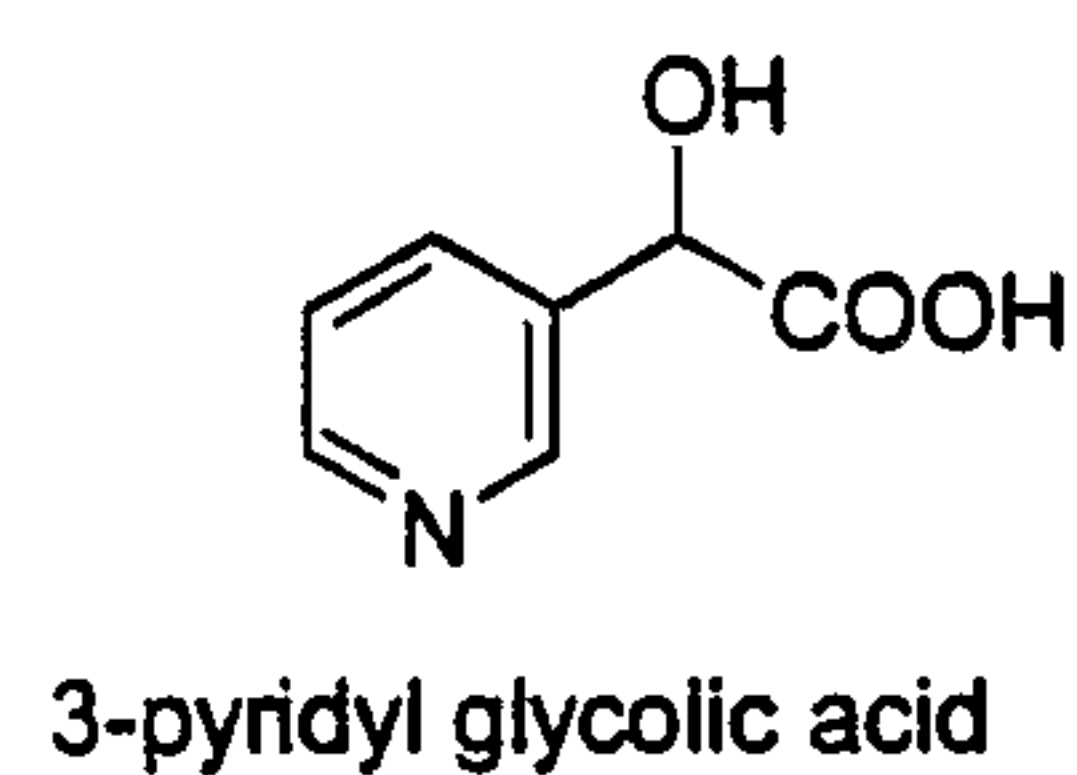
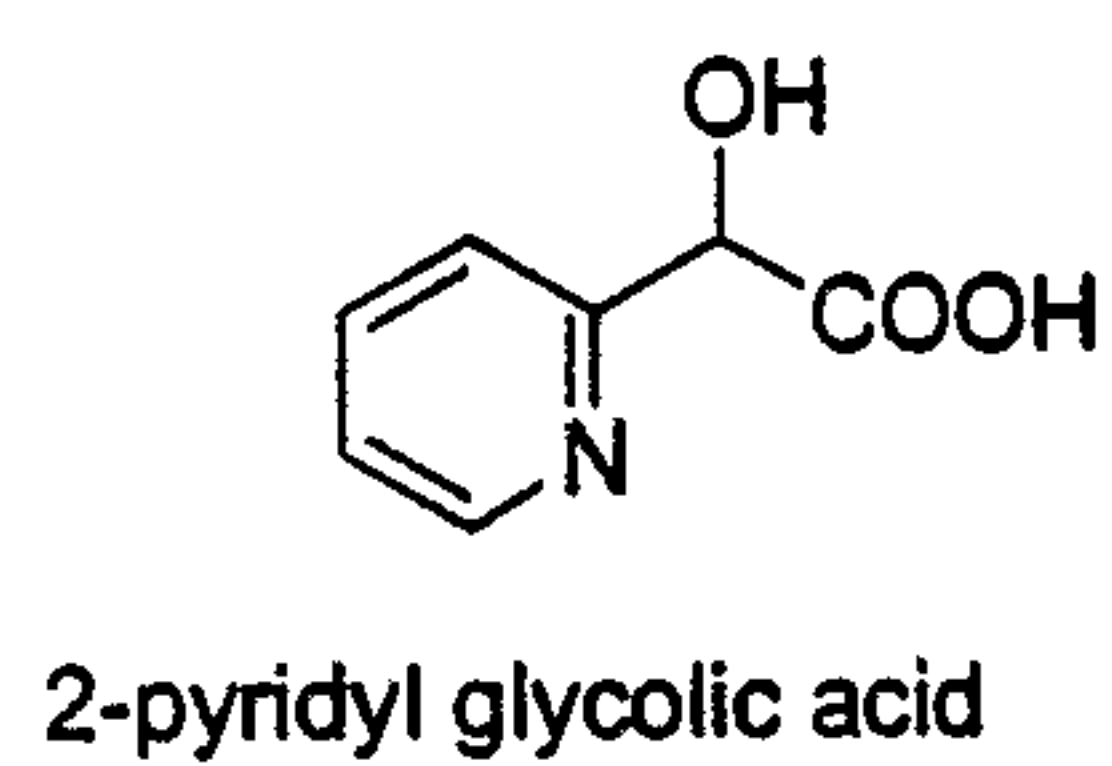
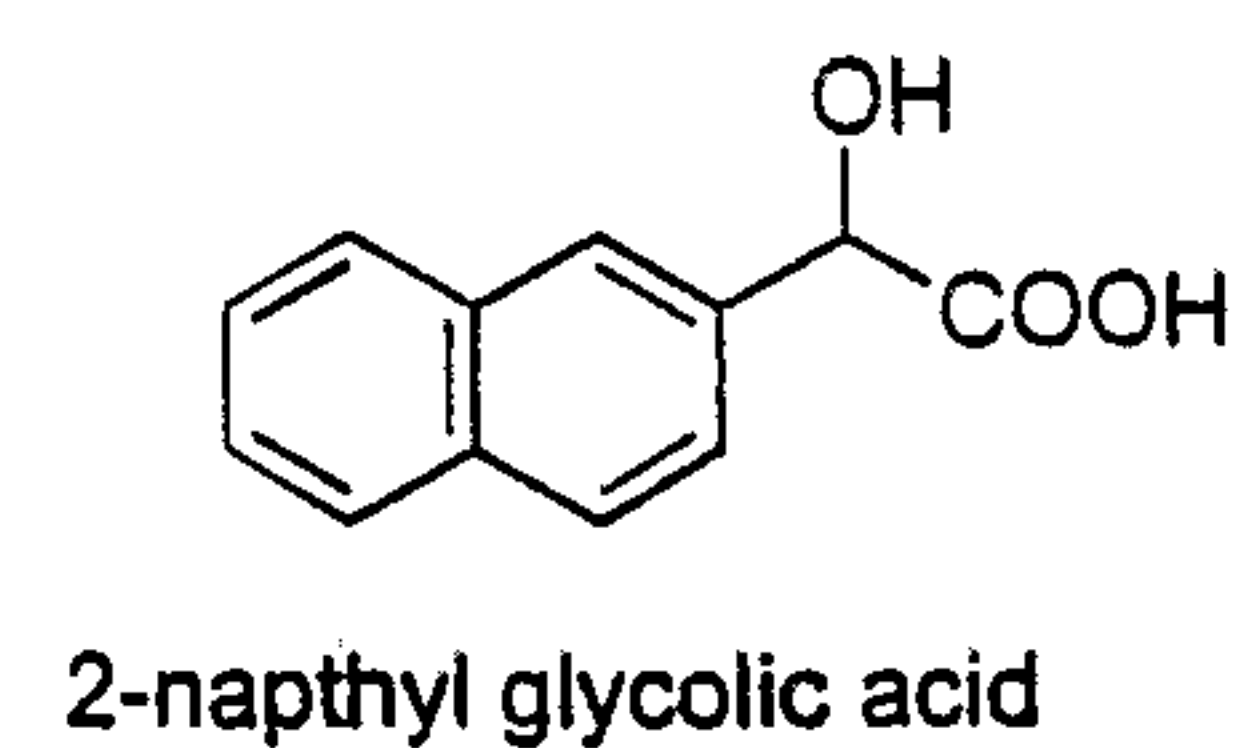
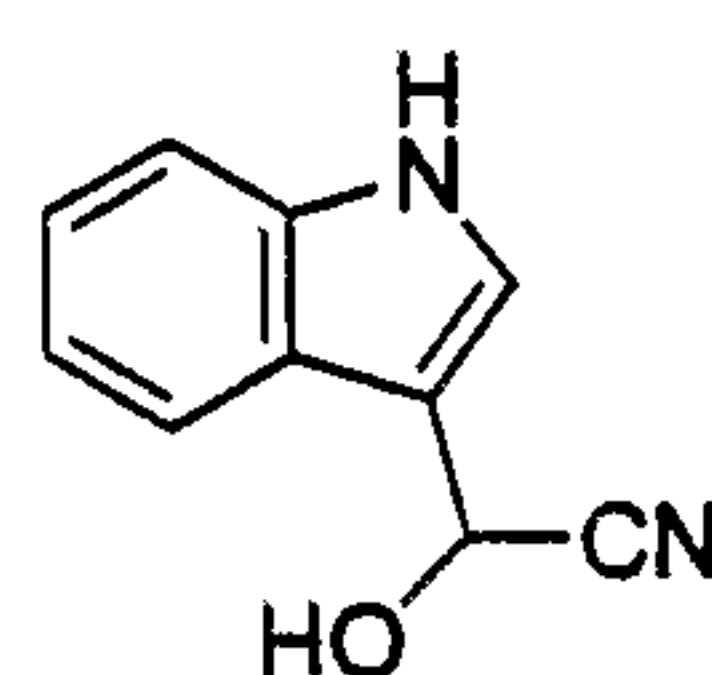
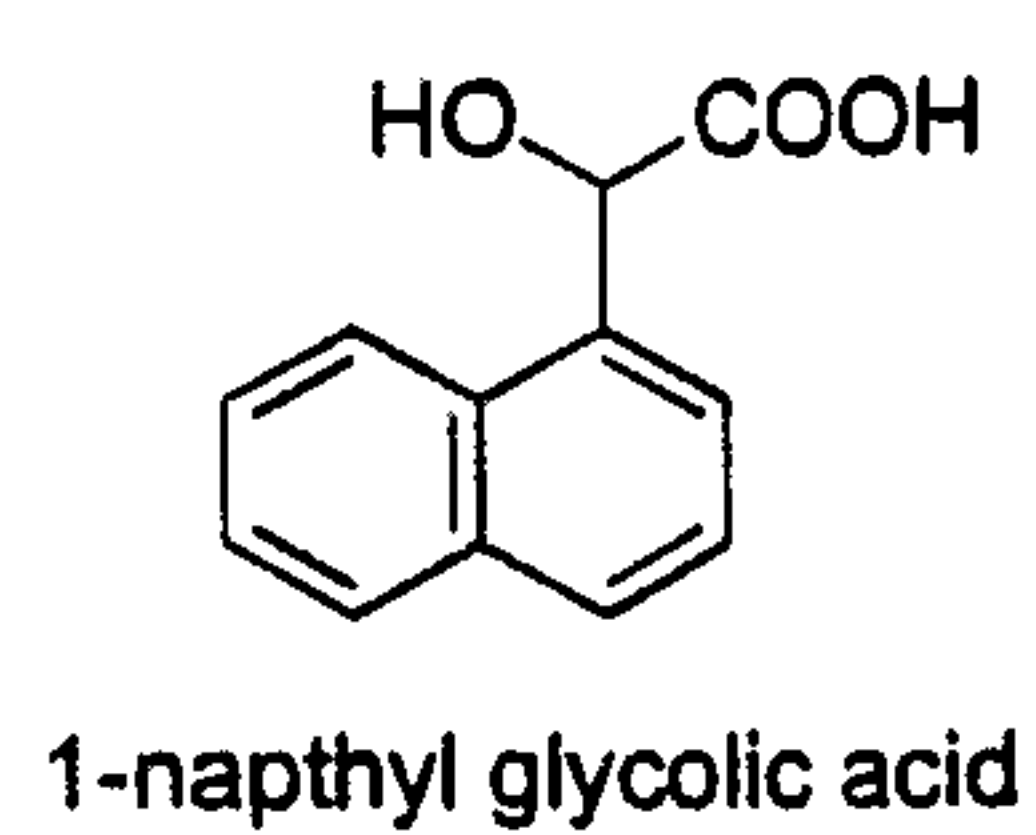
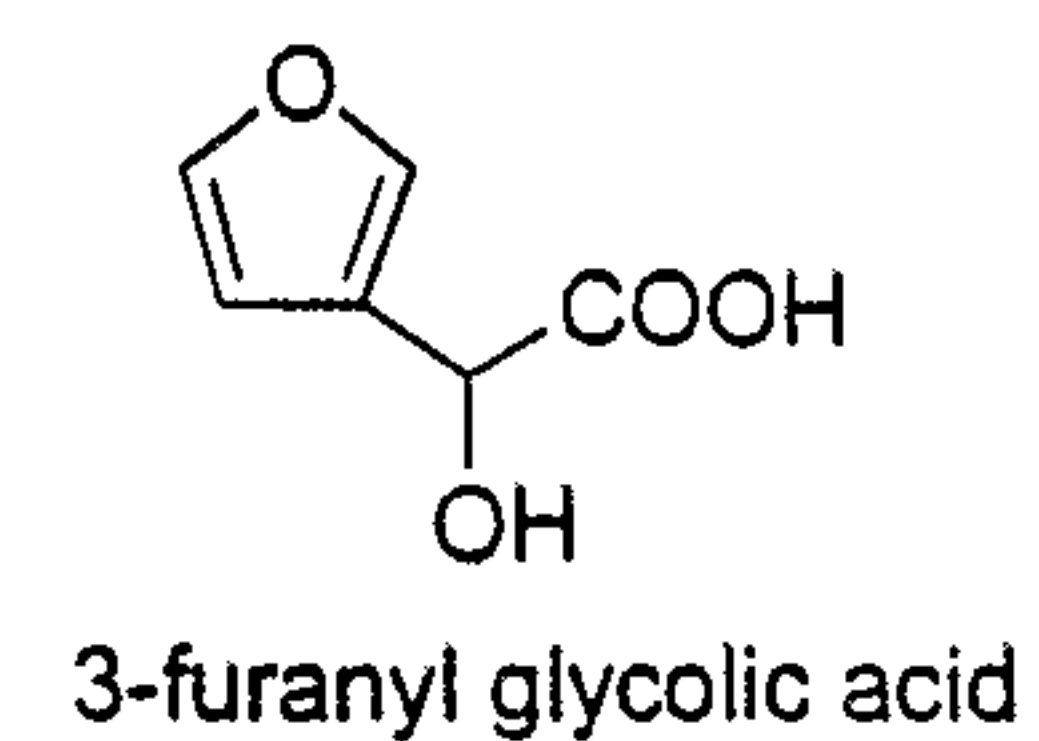
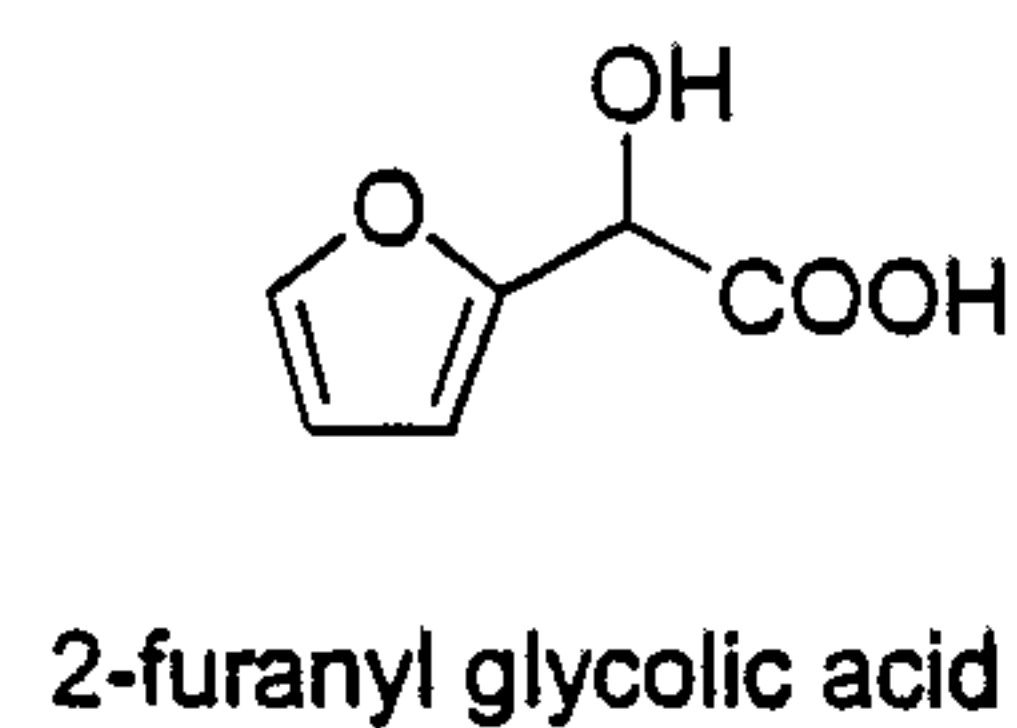
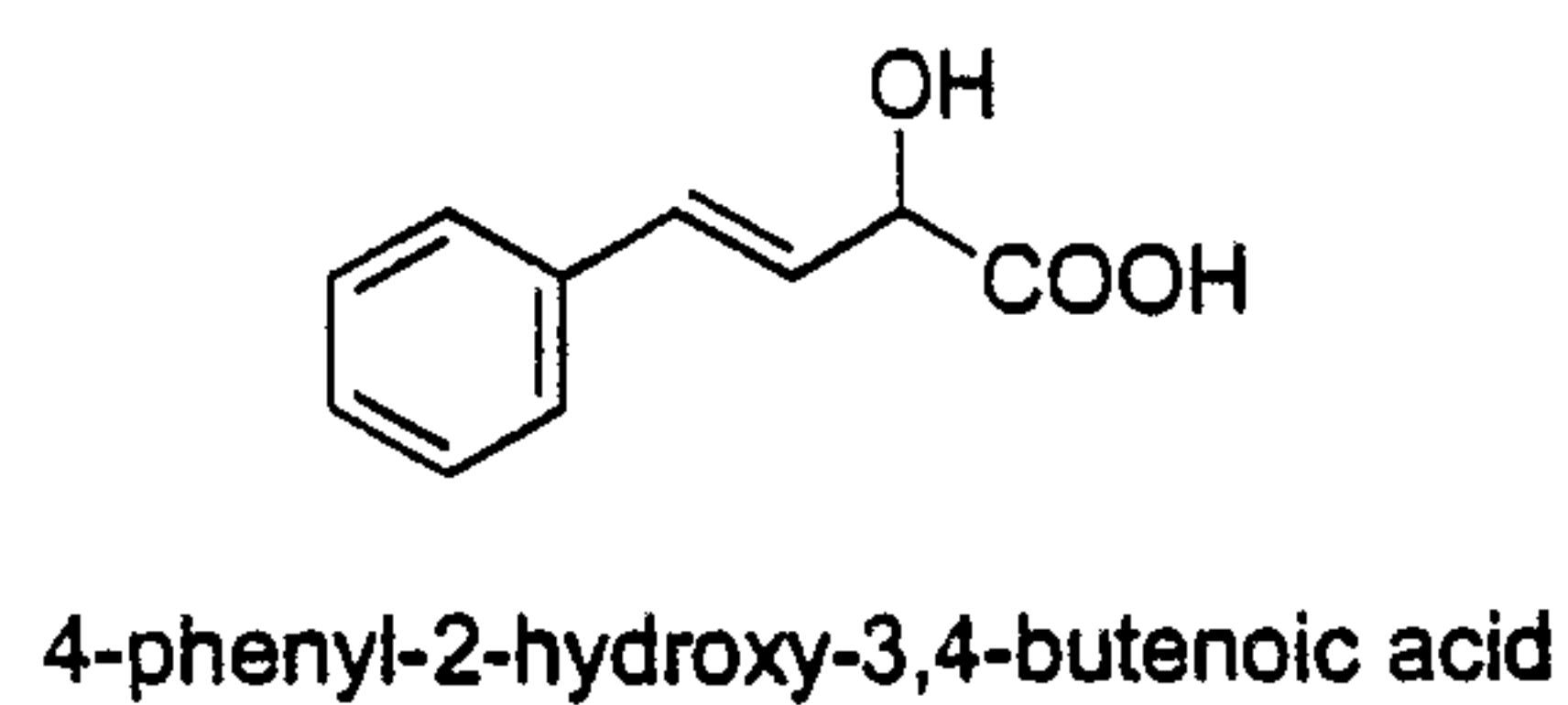
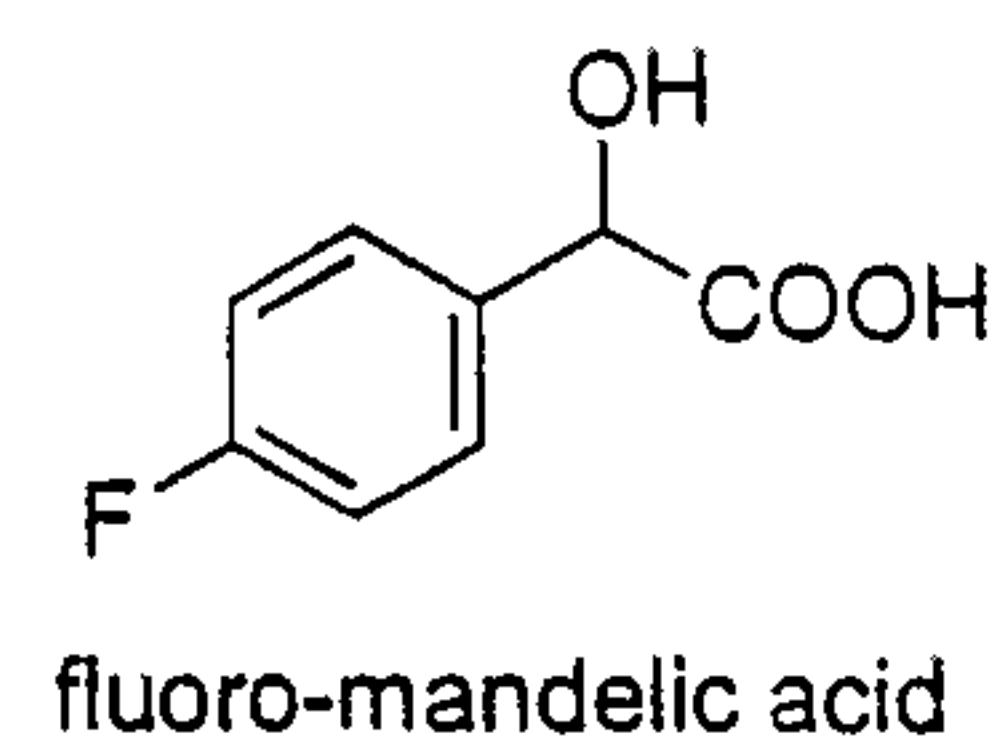
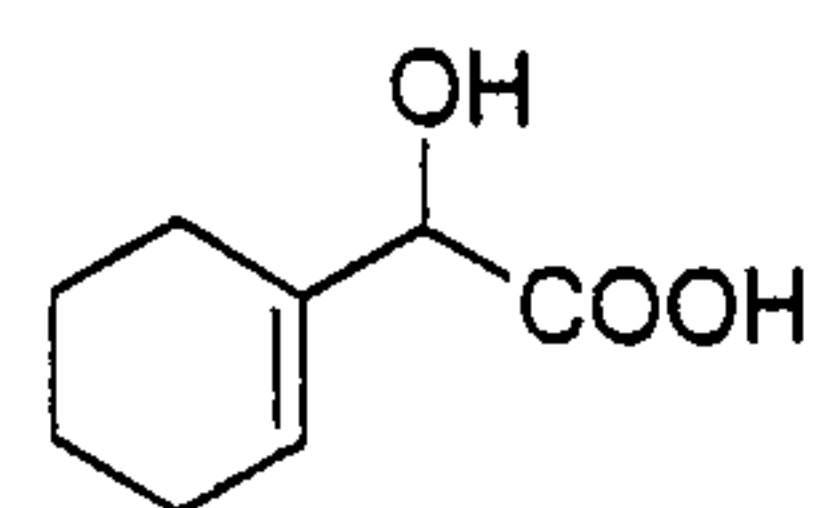
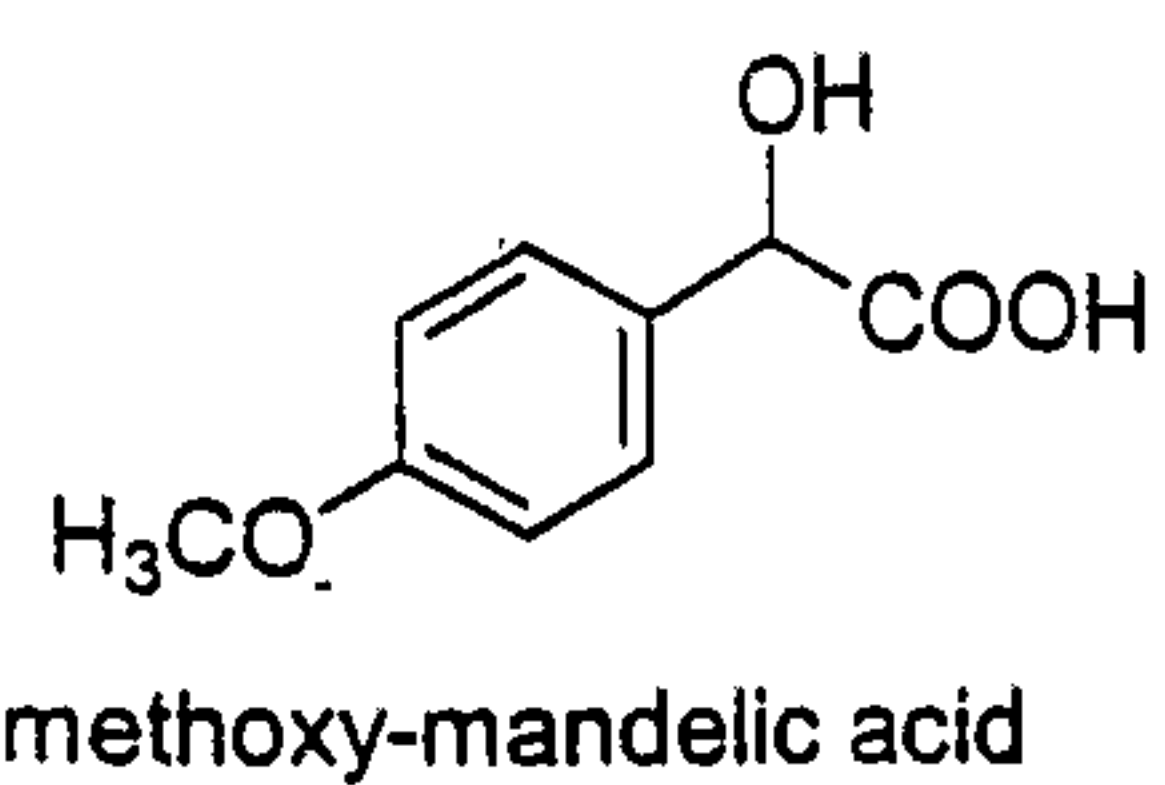




Figure 15

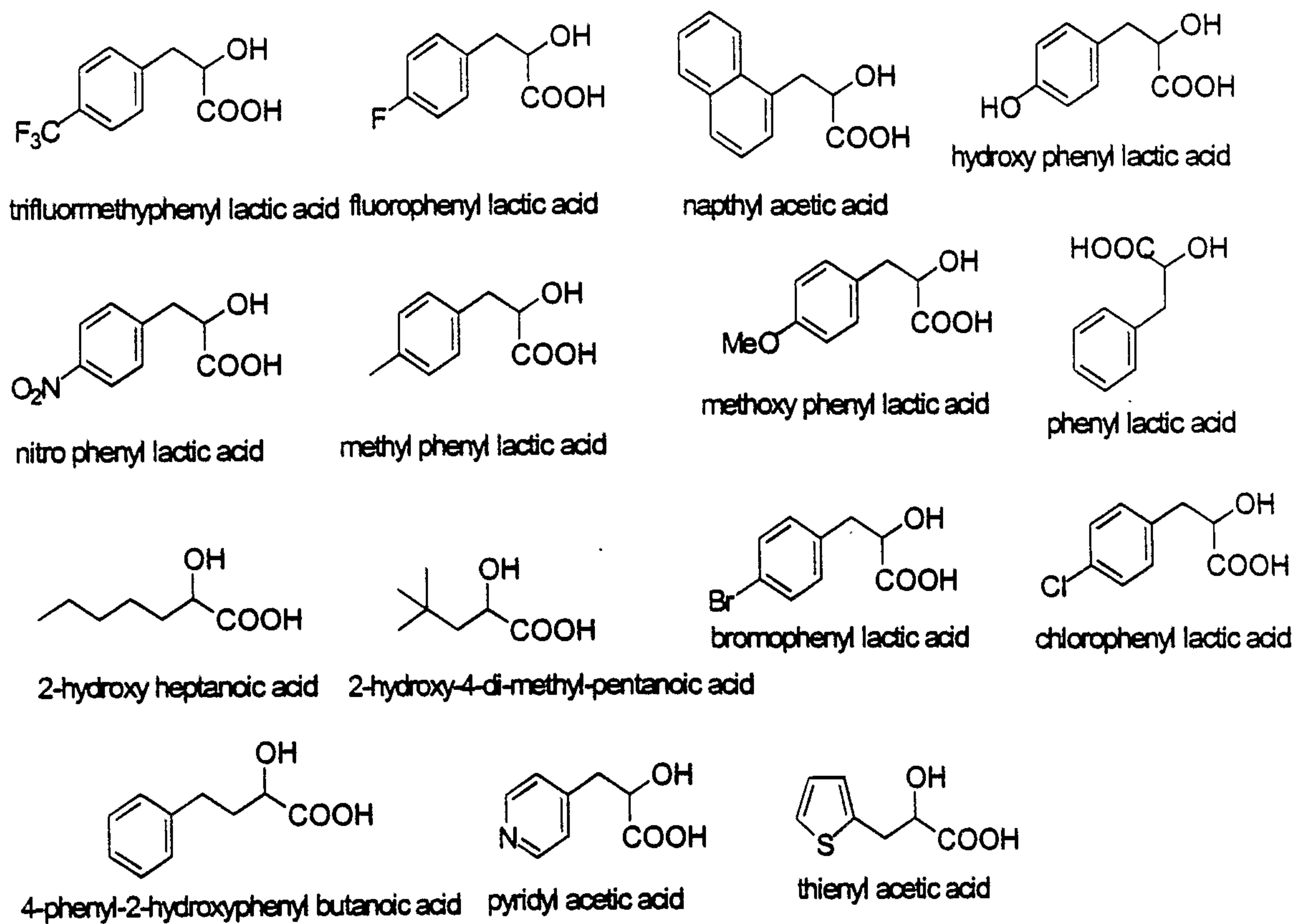




Figure 15

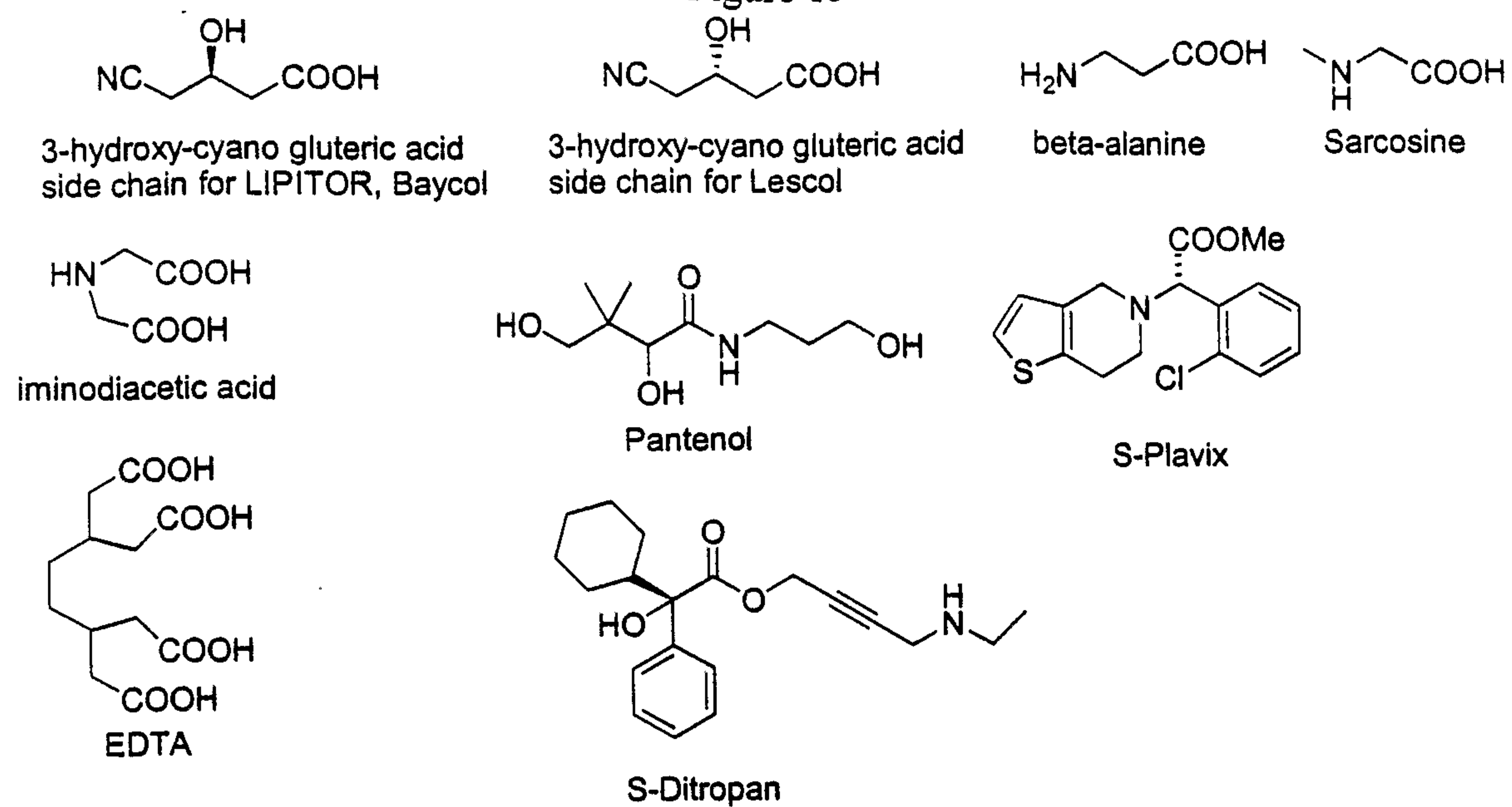
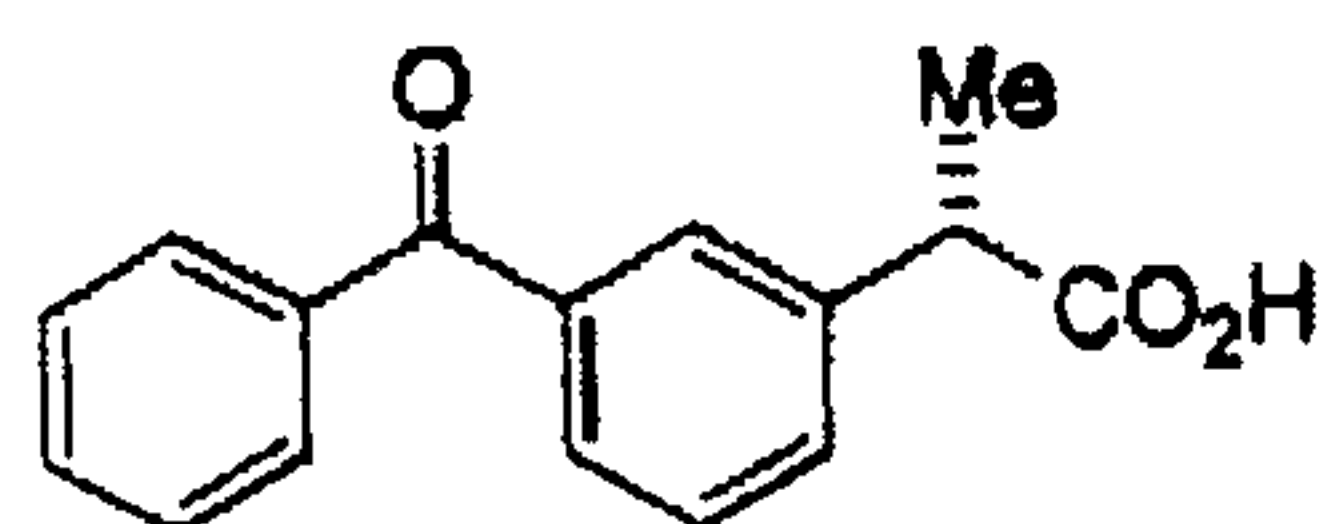
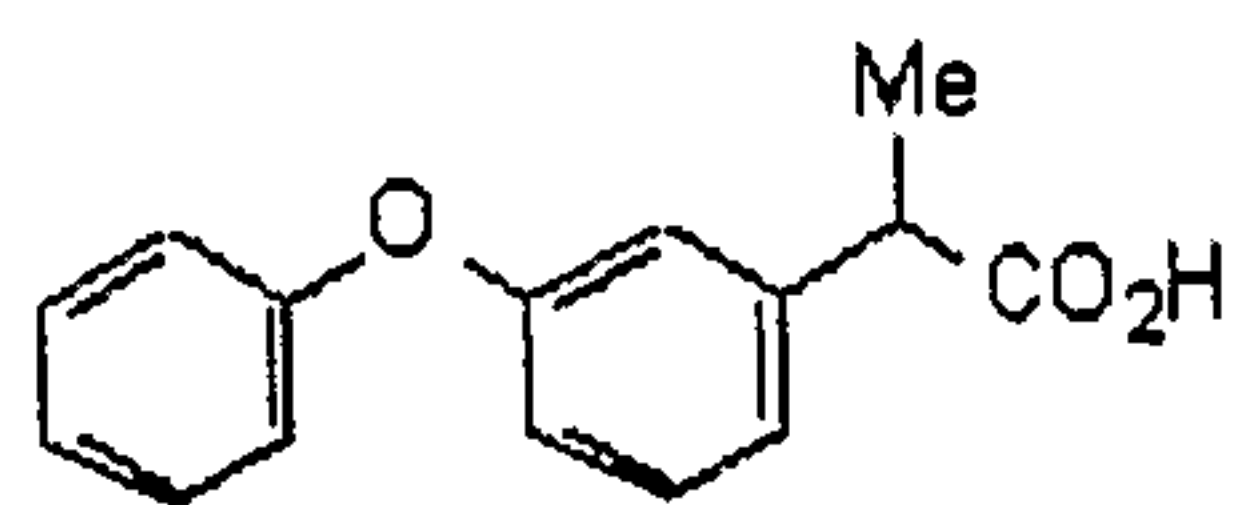




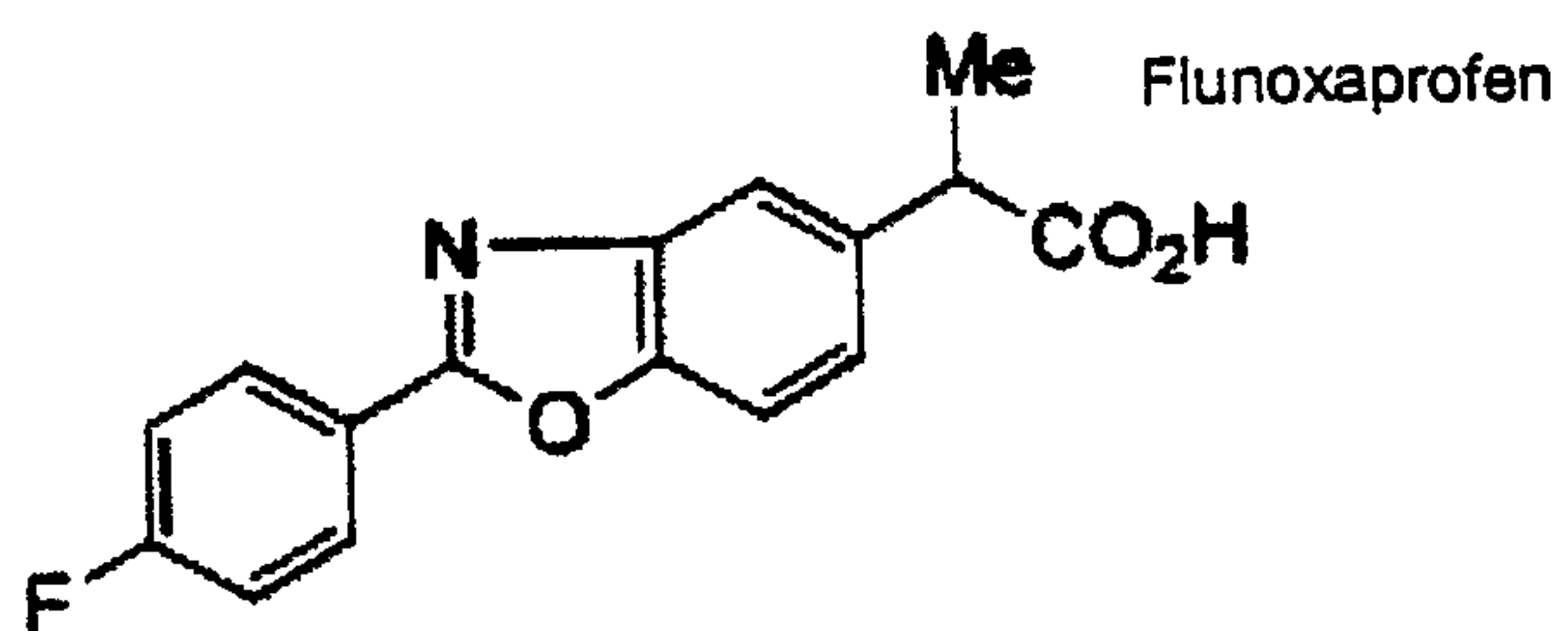
Figure 15



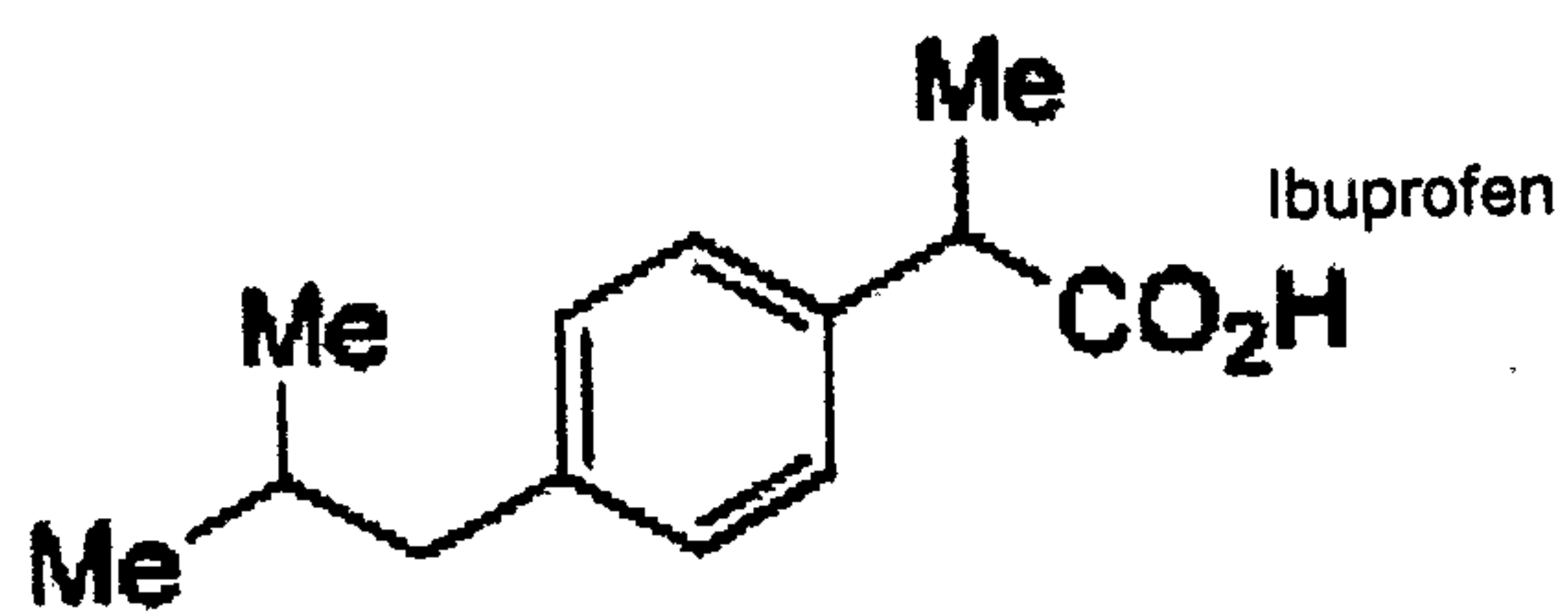
Dexketoprofen



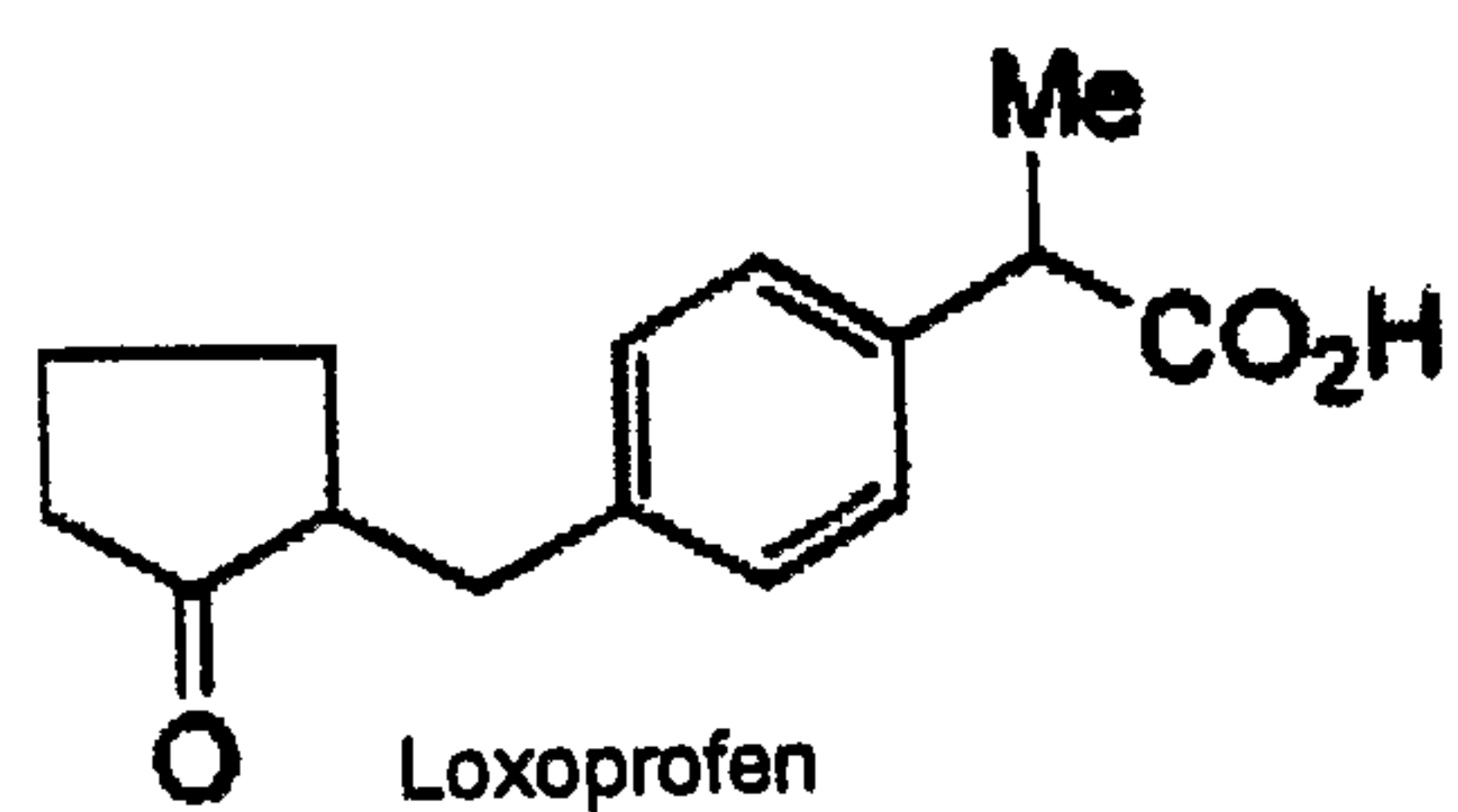
Fenoprofen



Flunoxaprofen



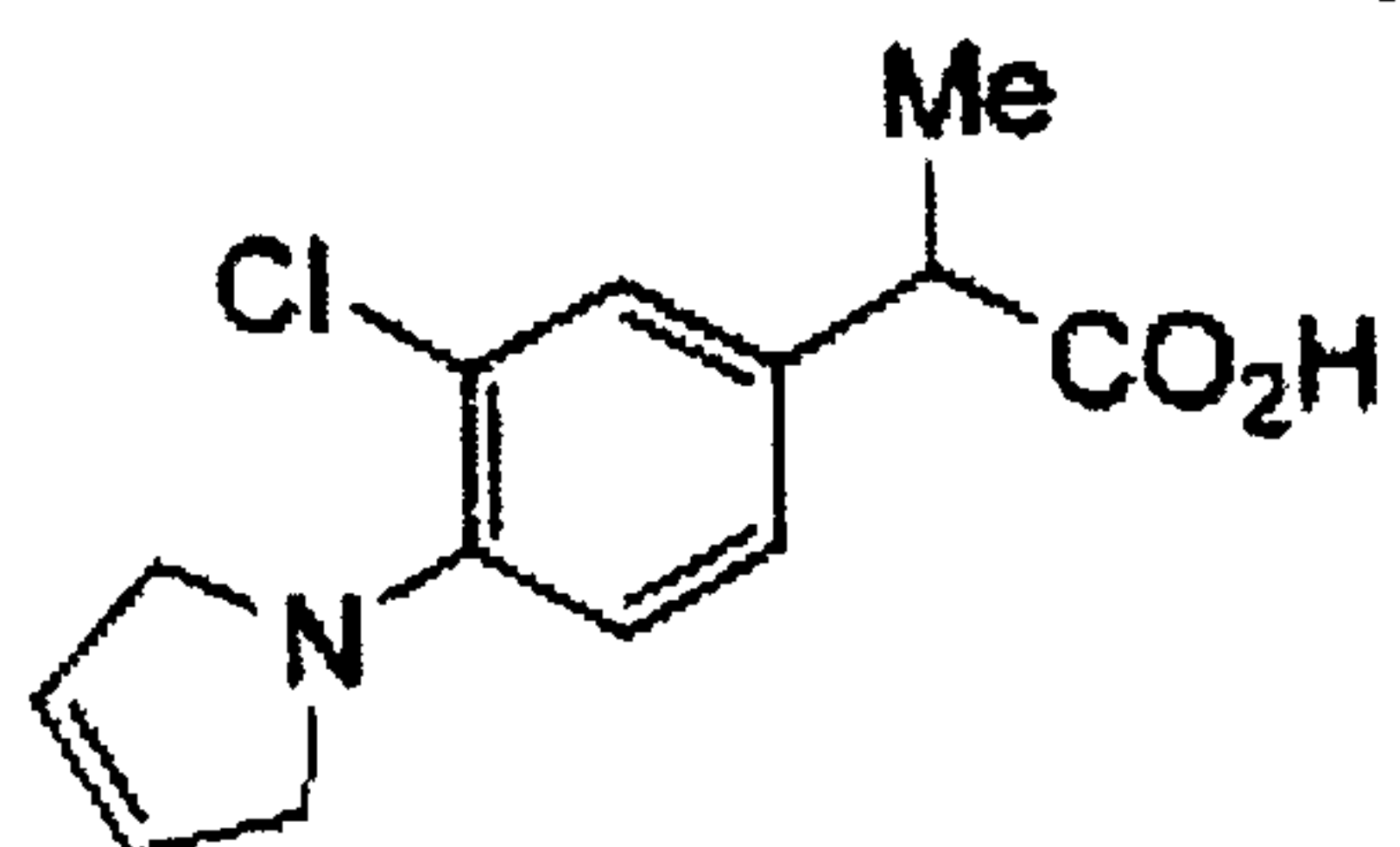
Ibuprofen



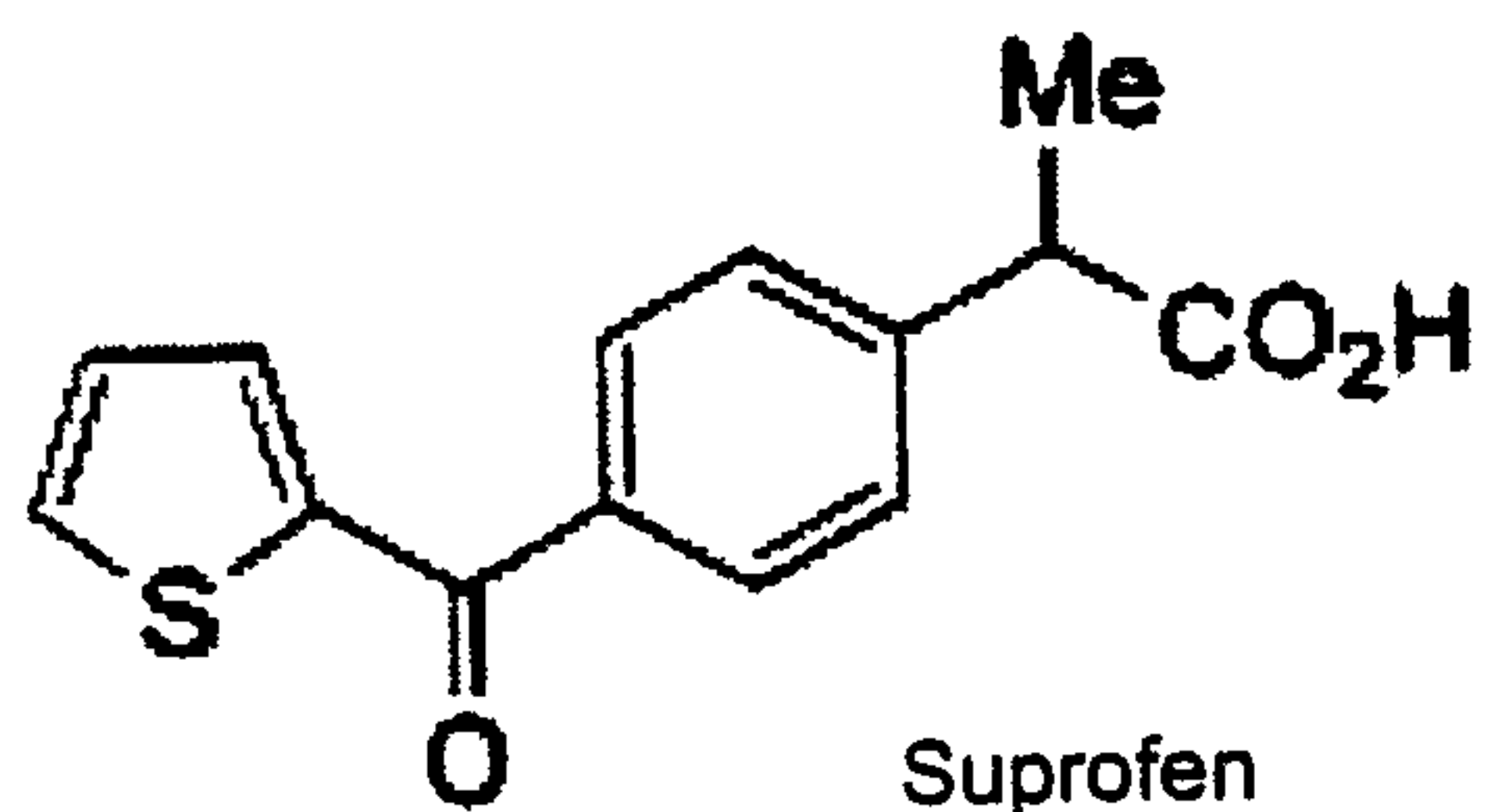
Loxoprofen



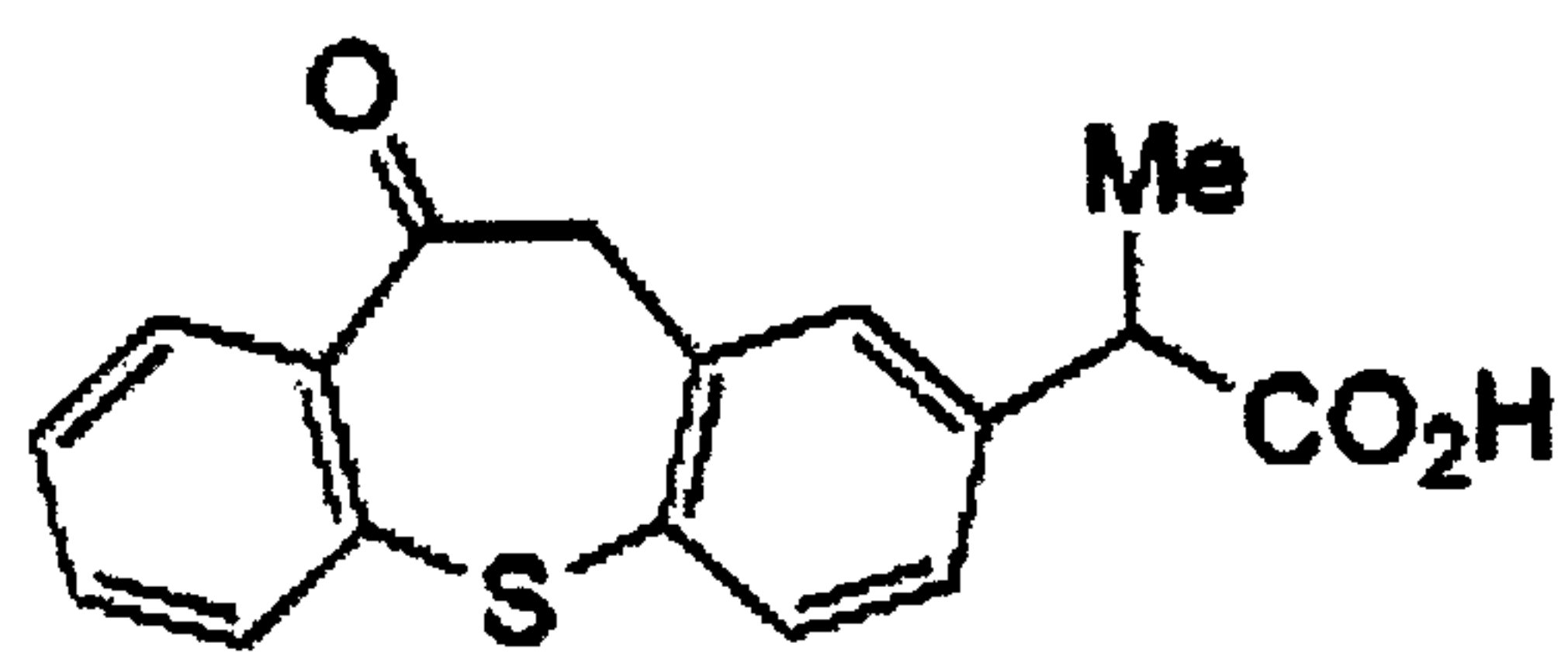
Figure 15



Pirprofen



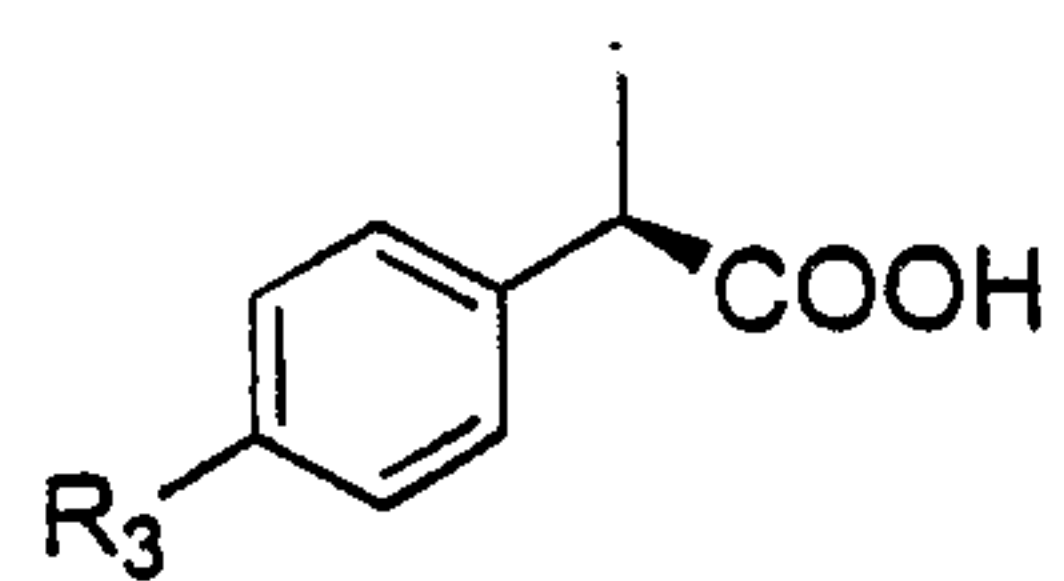
Suprofen



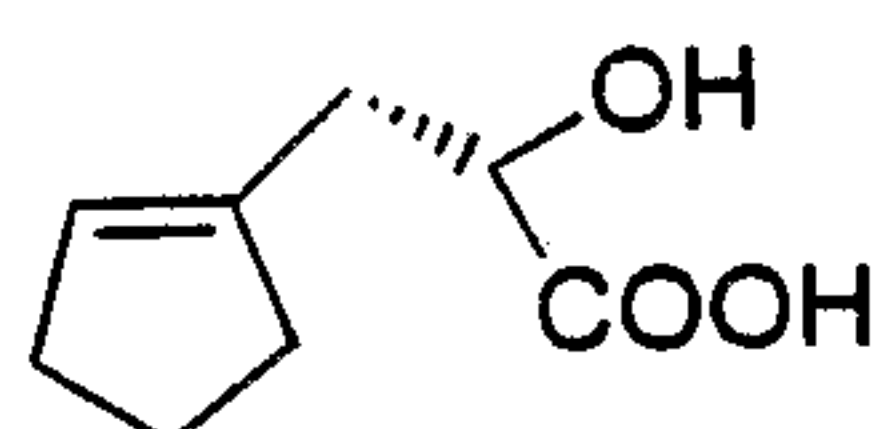
Zaltoprofen



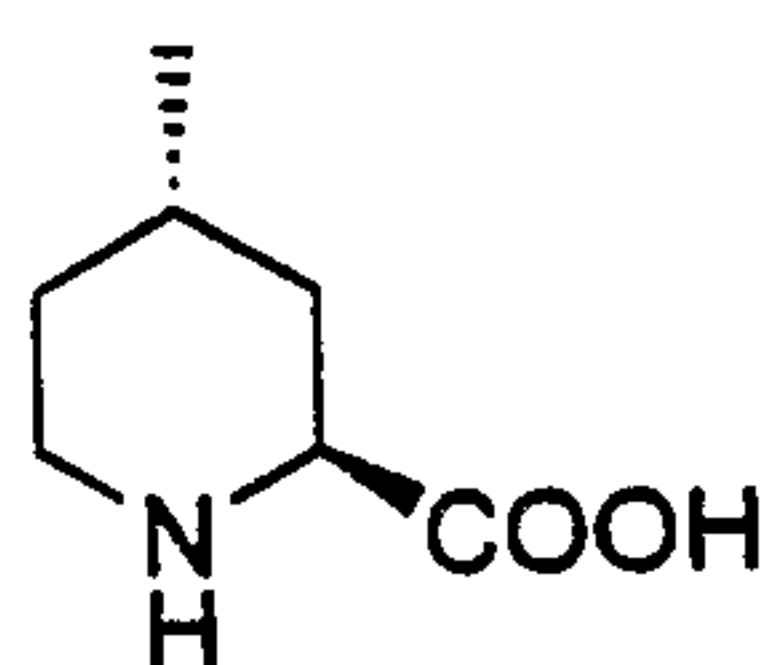
Figure 15



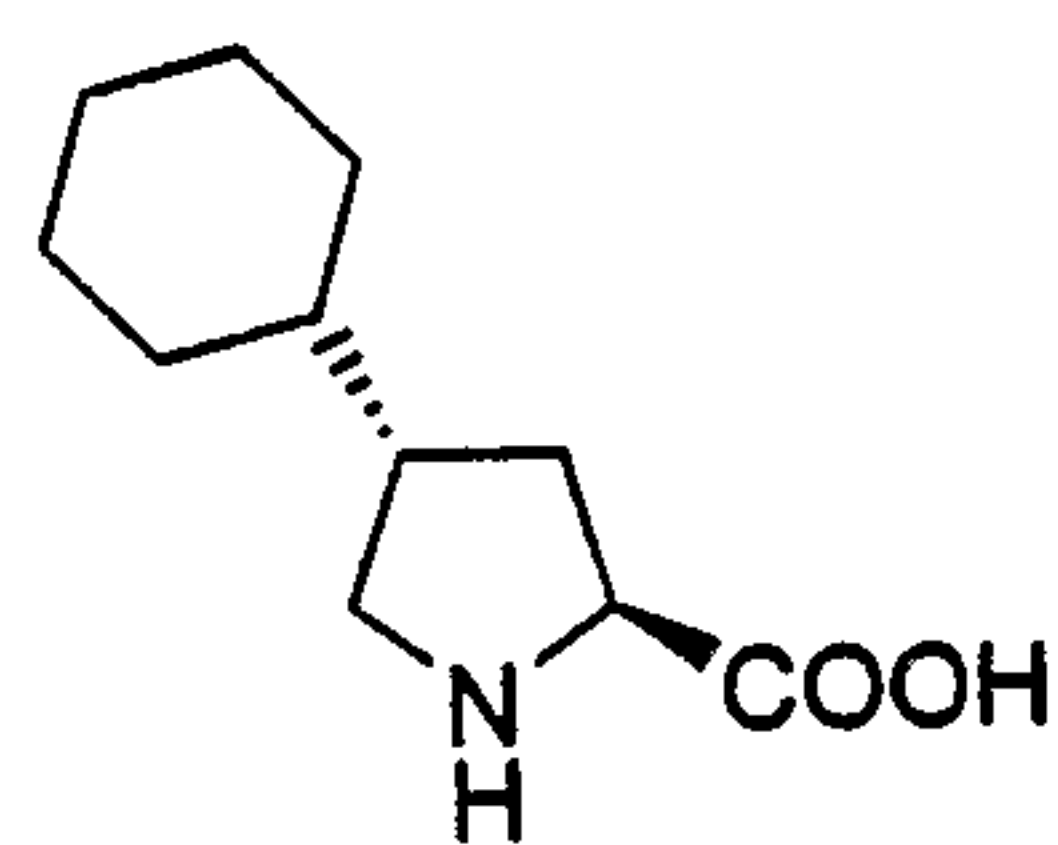
alpha-methyl benzyl cyanide derivatives



intermediate for Trocade



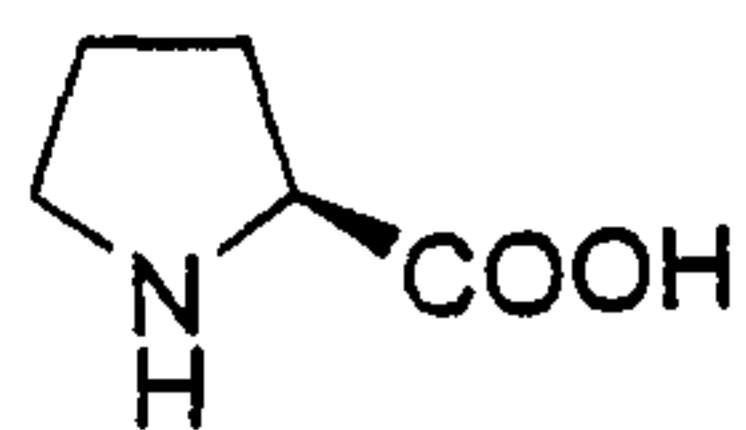
3-methyl-2-carboxy-piperidine



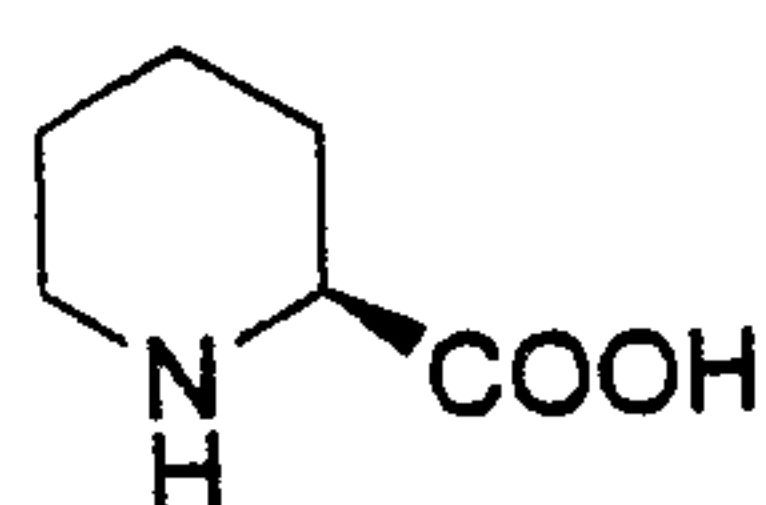
Fospril



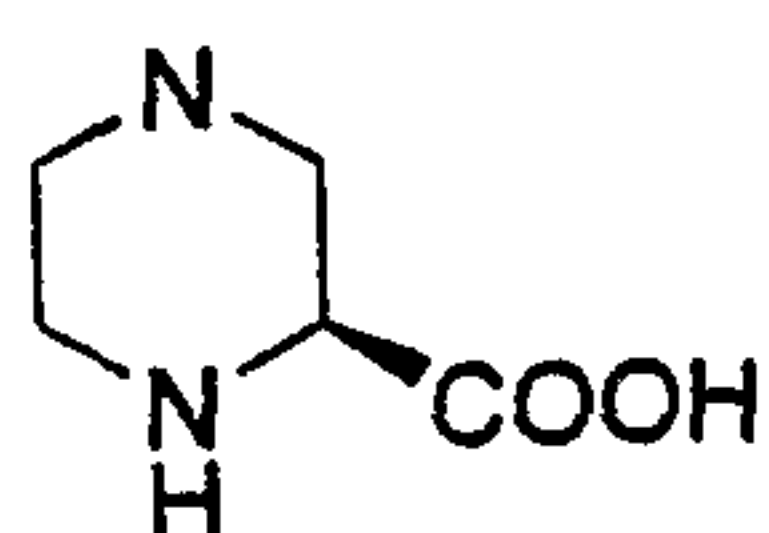
Figure 15



2-carboxy-cyclobutyl amine



2-carboxy-piperidine



2-carboxy-piperazine



Figure 16

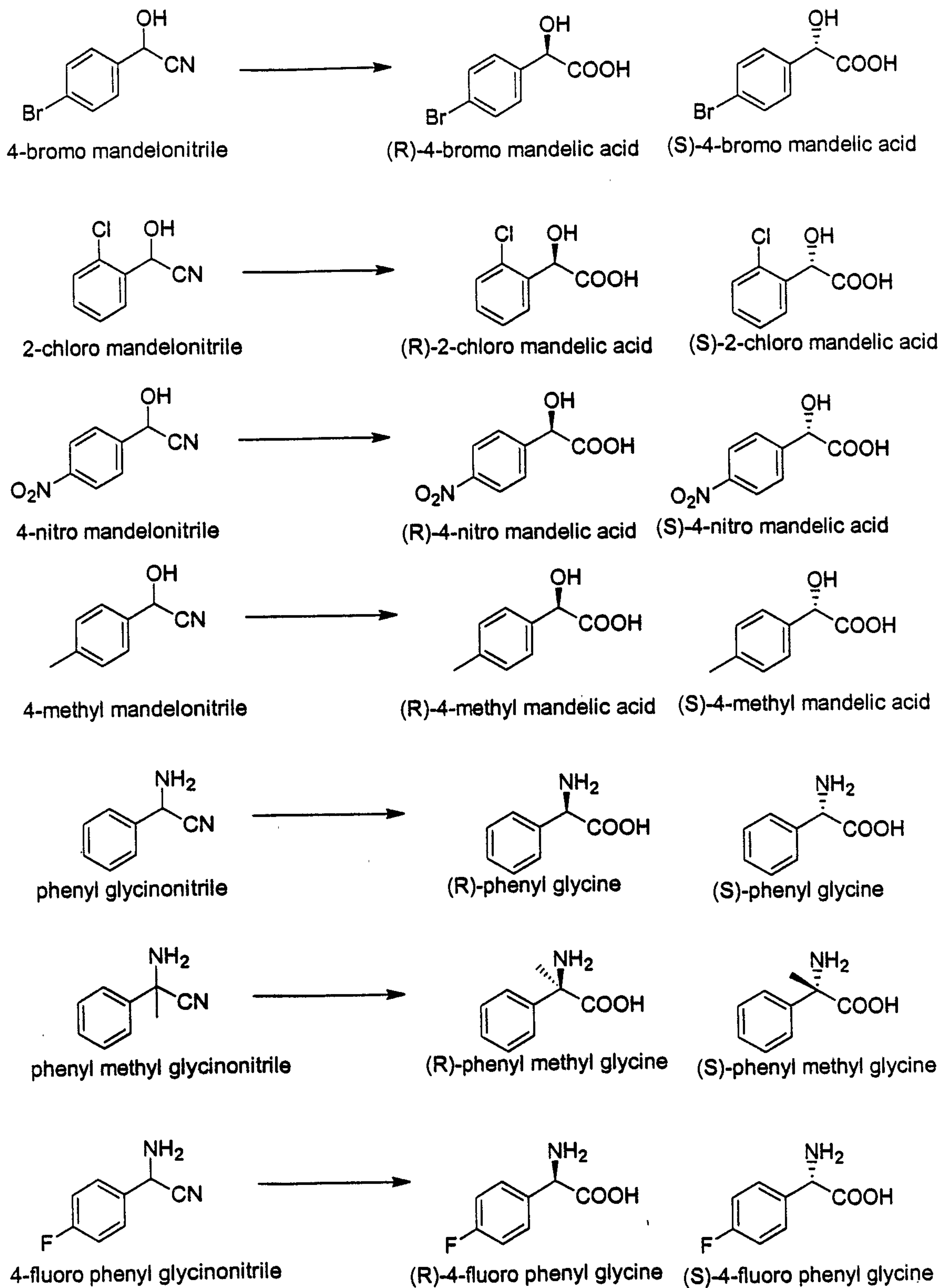




Figure 16

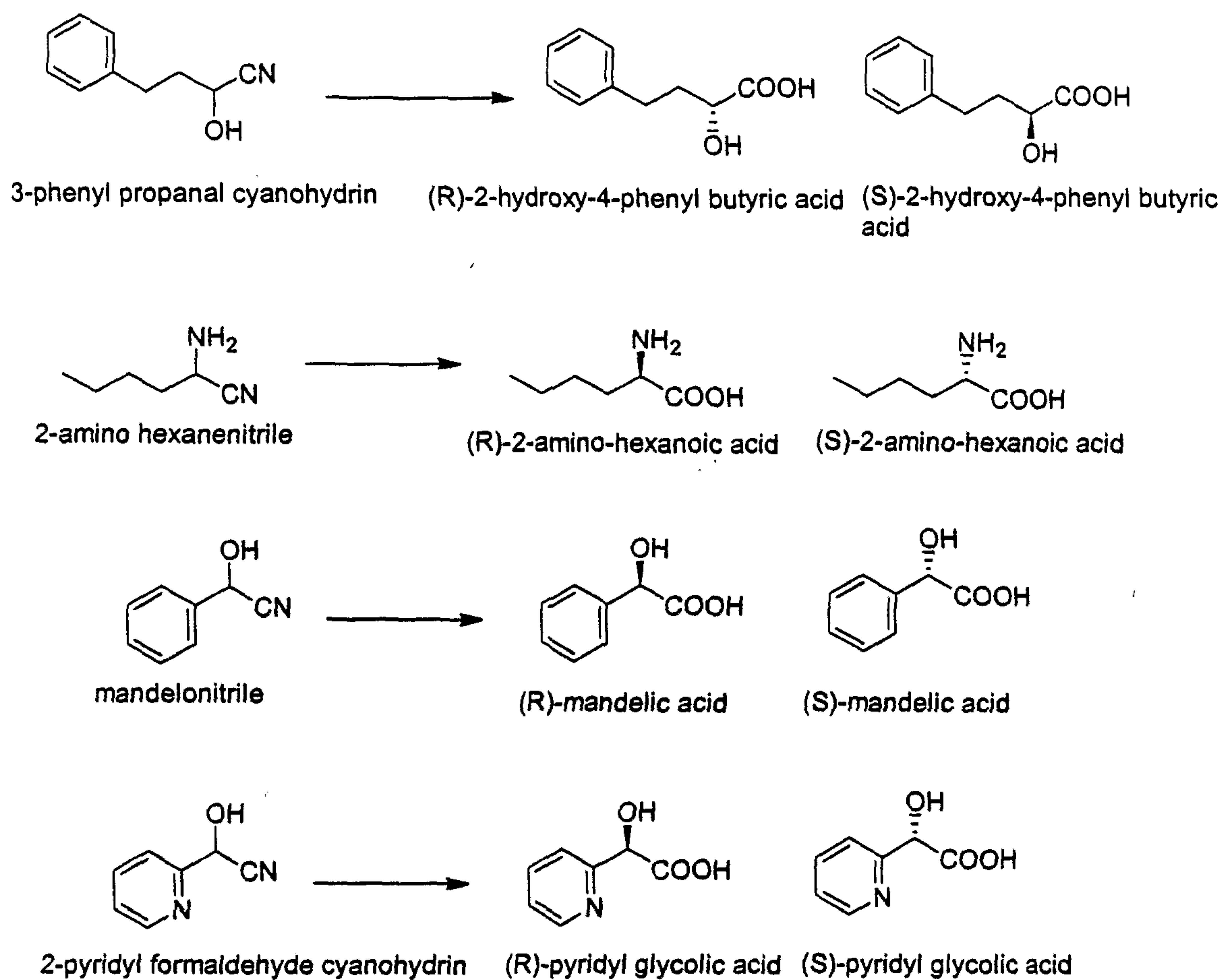




Figure 16

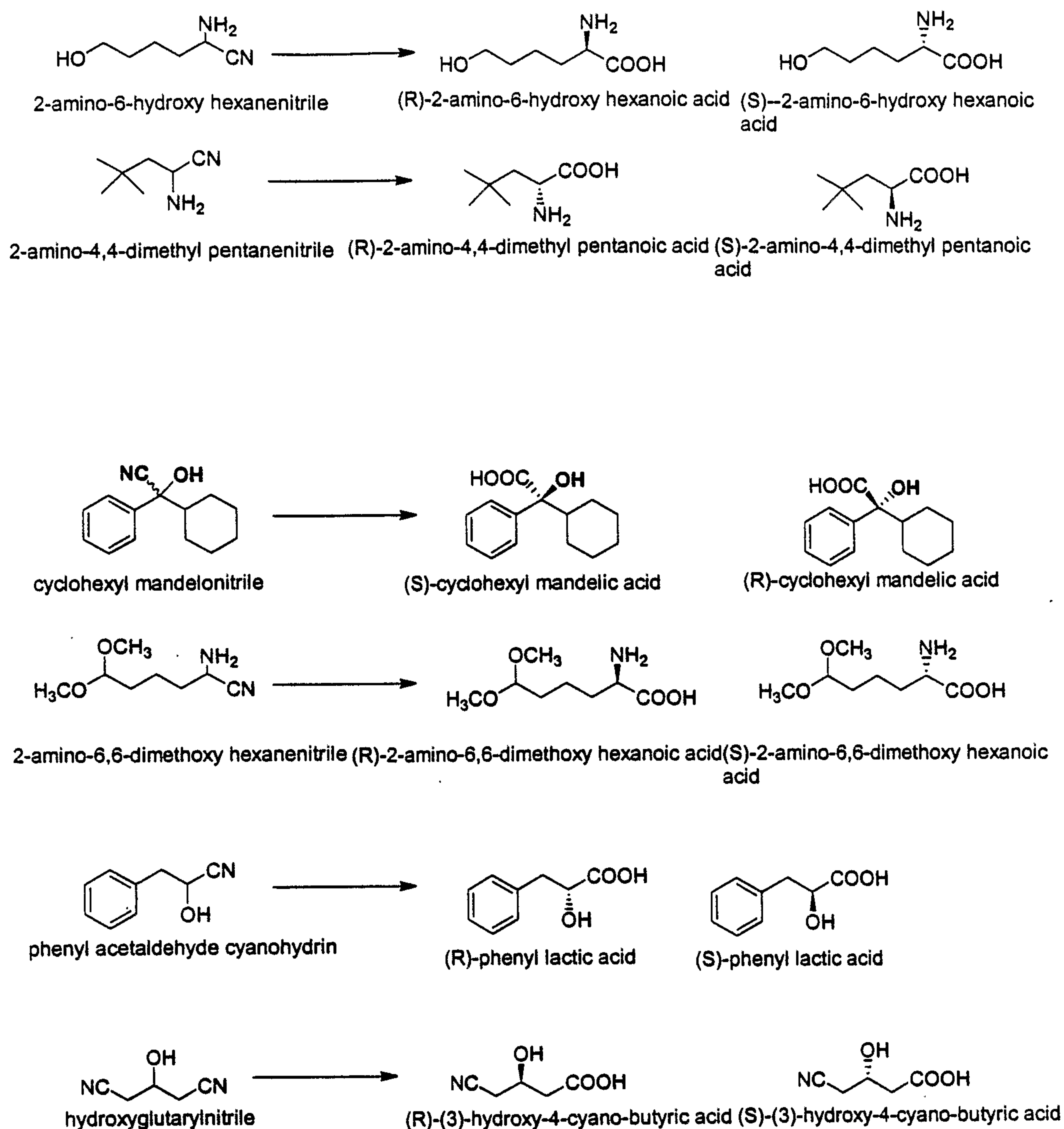




Figure 16

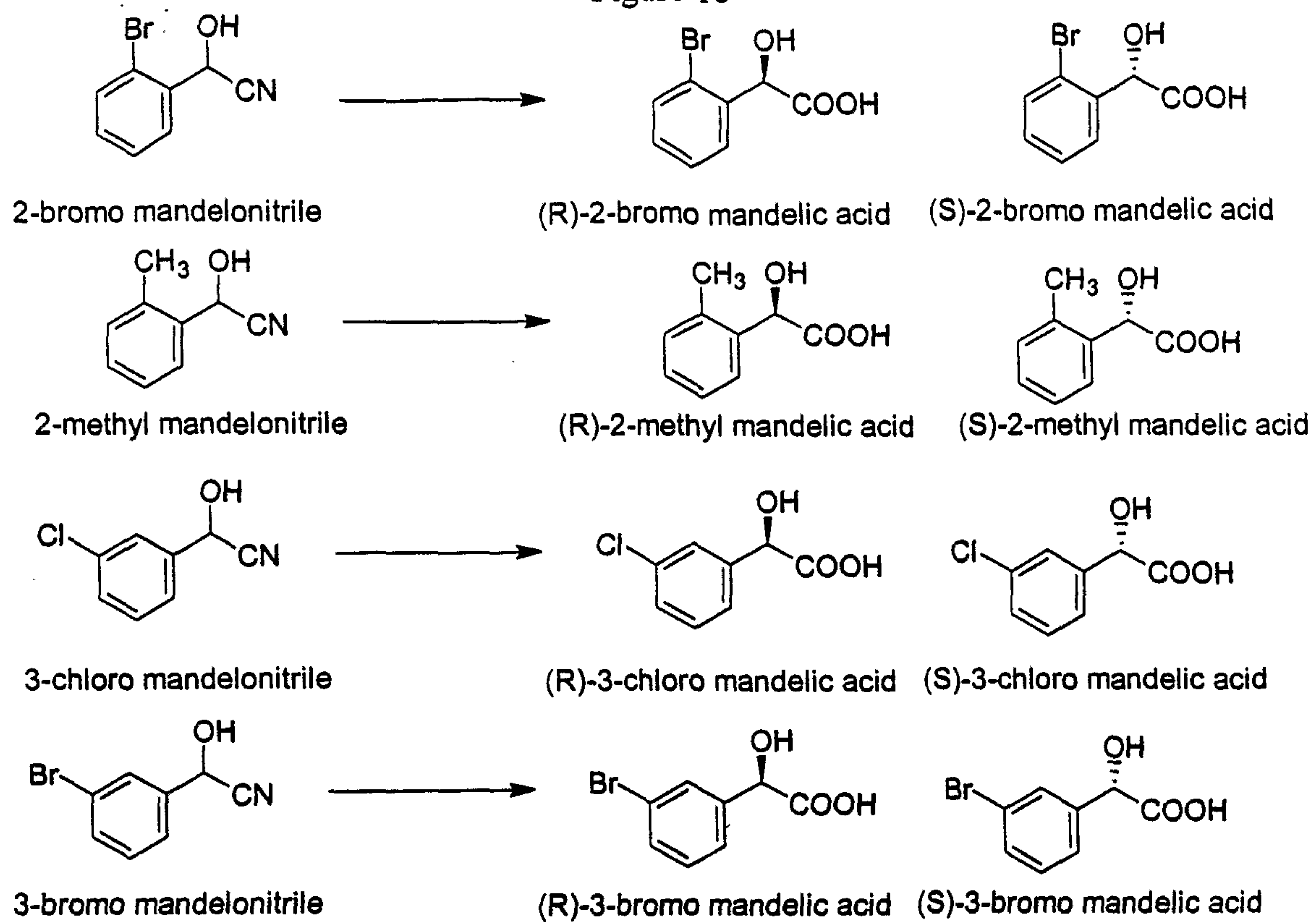




Figure 16

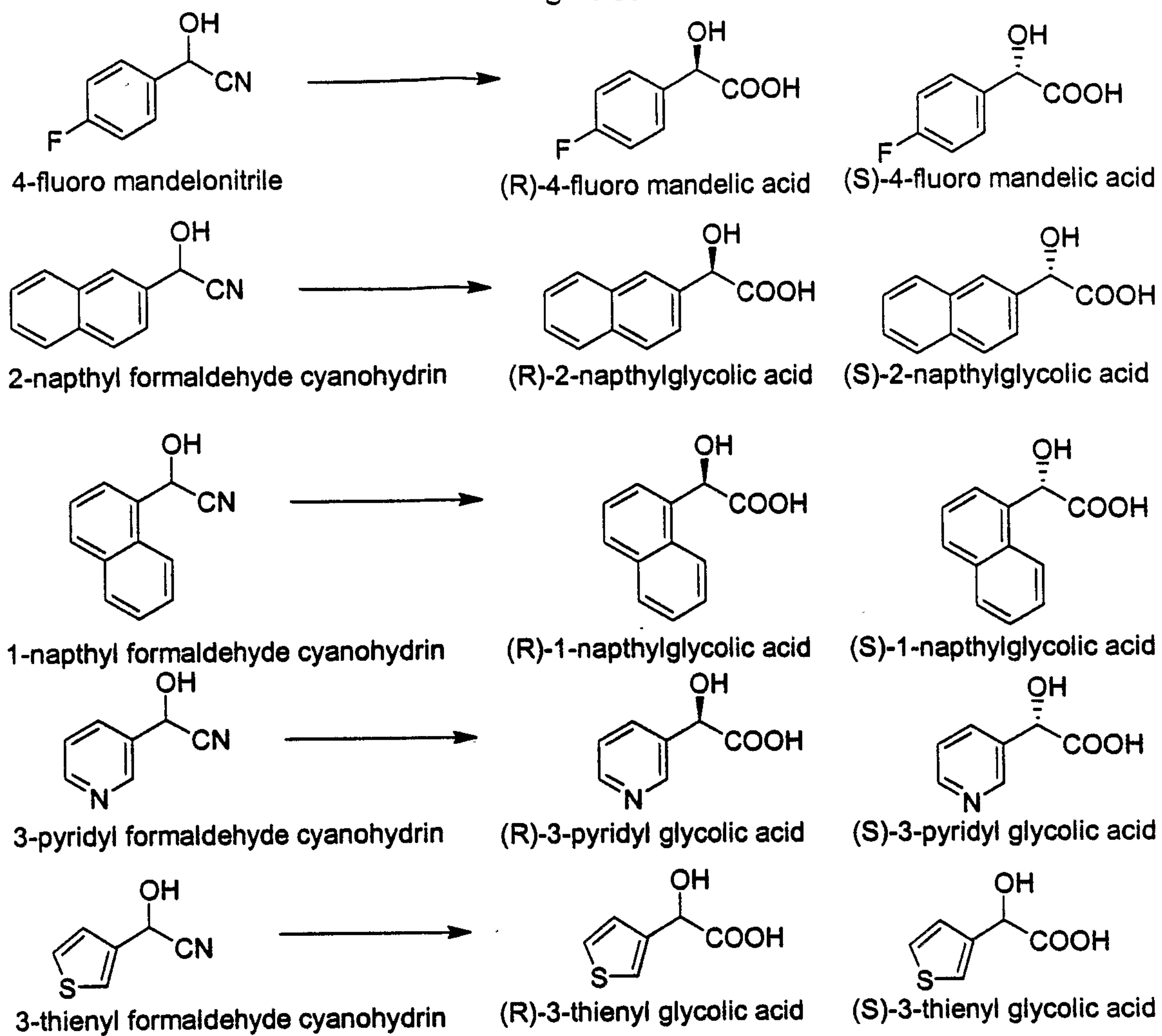




Figure 16

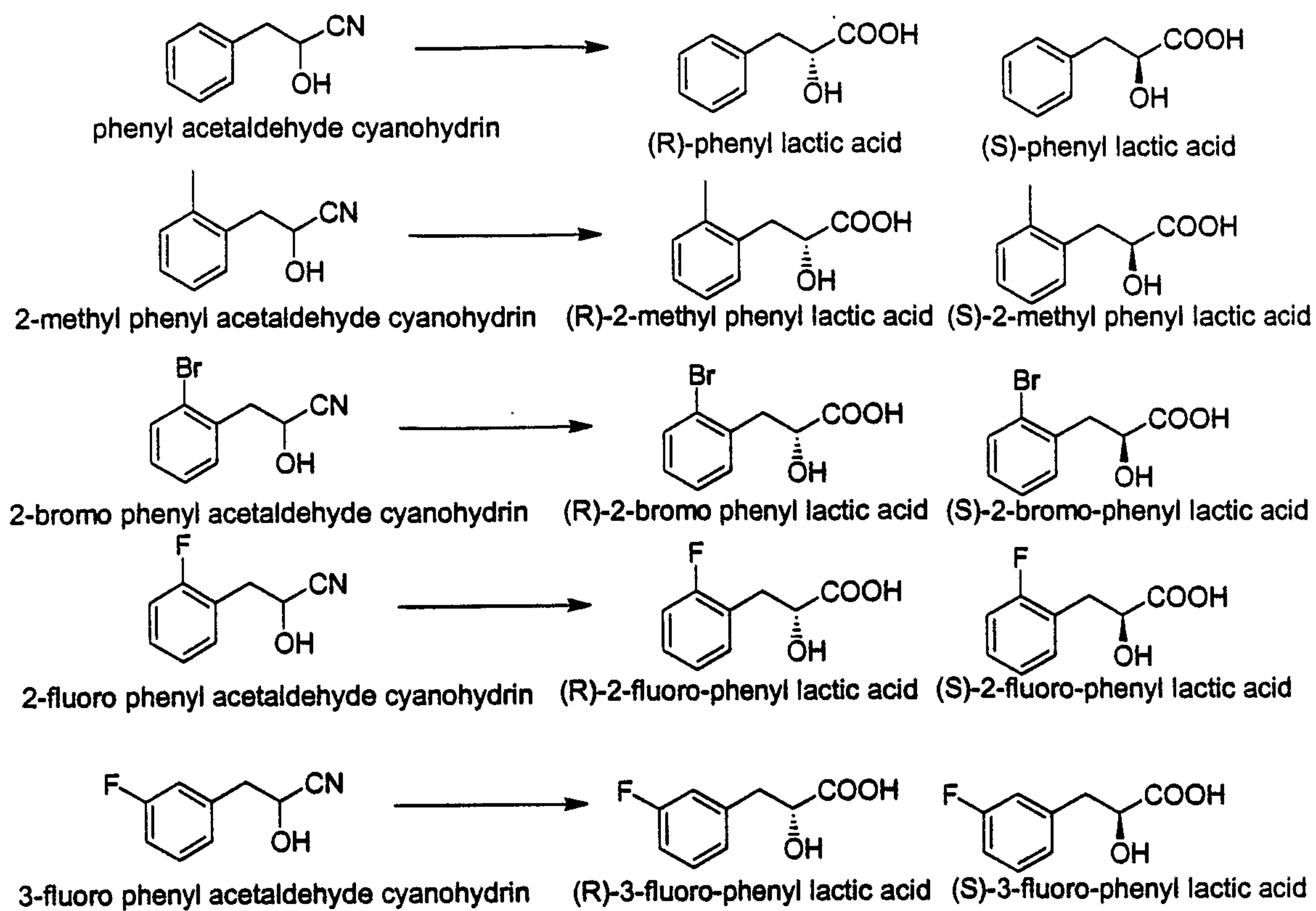




Figure 16

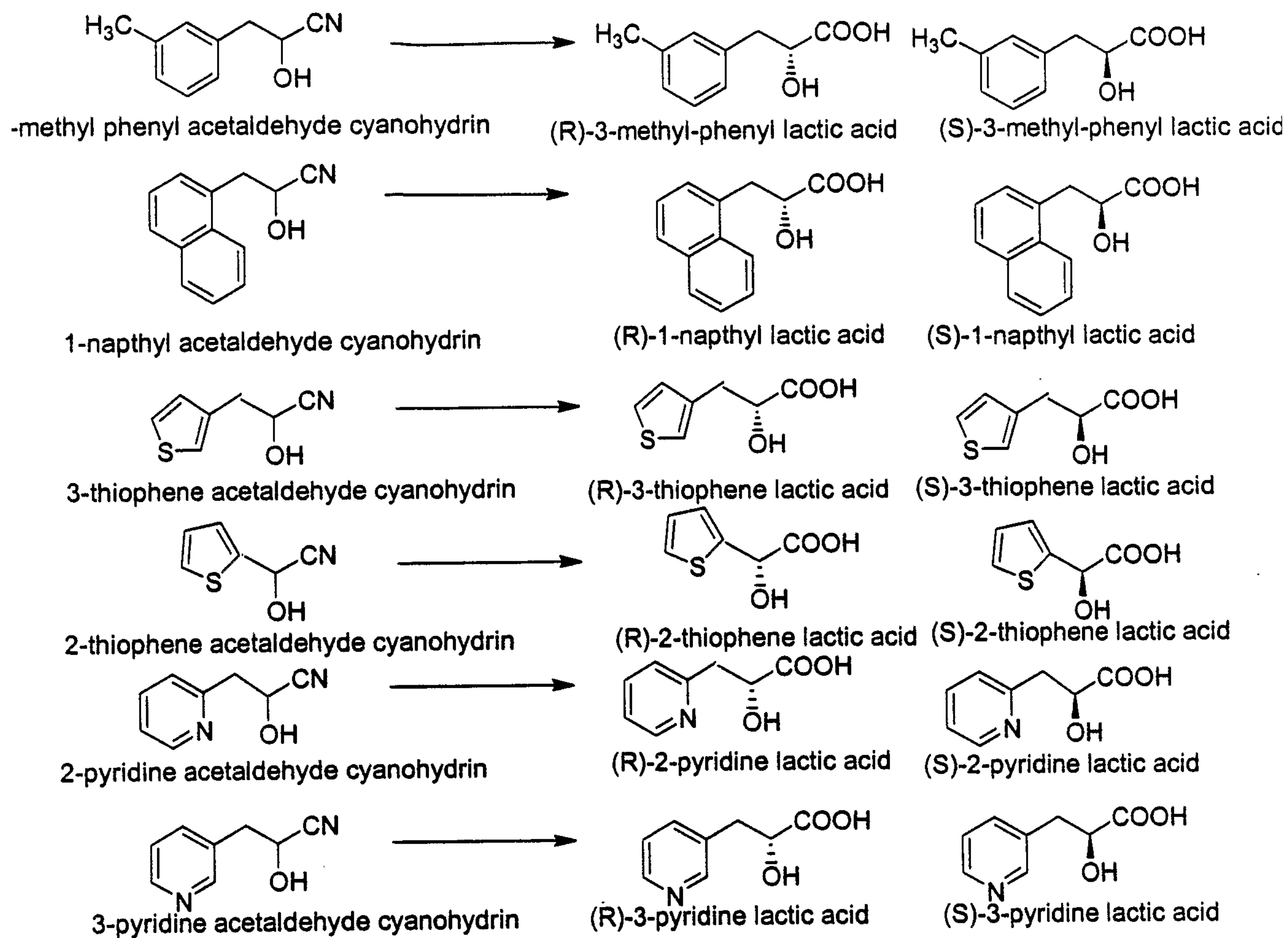




Figure 17

