(12) STANDARD PATENT (11) Application No. **AU 2008254837 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Chimeric PUFA polyketide synthase systems and uses thereof

(51) International Patent Classification(s)

A01H 5/00 (2006.01)	***C12N 1/15*** (2006.01)
A61K 48/00 (2006.01)	***C12N 1/19*** (2006.01)
A61P 1/04 (2006.01)	***C12N 1/20*** (2006.01)
A61P 1/16 (2006.01)	***C12N 1/21*** (2006.01)
A61P 3/06 (2006.01)	***C12N 5/04*** (2006.01)
A61P 9/10 (2006.01)	***C12N 5/10*** (2006.01)
A61P 11/00 (2006.01)	***C12N 9/02*** (2006.01)
A61P 15/06 (2006.01)	***C12N 15/00*** (2006.01)
A61P 19/02 (2006.01)	***C12N 15/09*** (2006.01)
A61P 19/10 (2006.01)	***C12N 15/11*** (2006.01)
A61P 25/08 (2006.01)	***C12N 15/52*** (2006.01)
A61P 25/24 (2006.01)	***C12N 15/82*** (2006.01)
A61P 25/28 (2006.01)	***C12P 1/02*** (2006.01)
A61P 29/00 (2006.01)	***C12P 7/62*** (2006.01)
A61P 31/04 (2006.01)	***C12P 7/64*** (2006.01)
A61P 35/00 (2006.01)	***C12P 21/04*** (2006.01)
A61P 37/02 (2006.01)	***C12Q 1/02*** (2006.01)
A61P 43/00 (2006.01)	***C12Q 1/68*** (2006.01)
C12N 1/00 (2006.01)	*C12R 1/645* (2006.01)
C12N 1/12 (2006.01)	

(21) Application No: **2008254837** (22) Date of Filing: **2008.05.16**

(87) WIPO No: **WO08/144473**

(30) Priority Data

(31) Number	(32) Date	(33) Country
11/749,686	**2007.05.16**	**US**

(43) Publication Date: **2008.11.27**
(44) Accepted Journal Date: **2013.09.12**

(71) Applicant(s)
DSM IP Assets B.V.

(72) Inventor(s)
Weaver, Craig A.;Doherty, Daniel H.;Zirkle, Ross;Metz, James G.

(74) Agent / Attorney
Allens Patent & Trade Mark Attorneys, Deutsche Bank Place Corner Hunter and Phillip Streets, SYDNEY, NSW, 2000

(56) Related Art
US 20050100995

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

(19) World Intellectual Property Organization
International Bureau

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

PCT

(10) International Publication Number
WO 2008/144473 A3

(51) International Patent Classification:
C12P 7/64 (2006.01)

(21) International Application Number:
PCT/US2008/063835

(22) International Filing Date: 16 May 2008 (16.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/749,686 16 May 2007 (16.05.2007) US

(71) Applicant *(for all designated States except US)*: **MARTEK BIOSCIENCES CORPORATION** [US/US]; 6480 Dobbin Road, Columbia, MD 21045 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WEAVER, Craig, A.** [US/US]; 4542 Starboard Drive, Boulder, CO 80301 (US). **ZIRKLE, Ross** [US/US]; 974 Wolf Creek Drive, Longmont, CO 80501 (US). **DOHERTY, Daniel, H.** [US/US]; 719 Ithaca Drive, Boulder, CO 80305 (US). **METZ, James, G.** [US/US]; 542 E. 16th Avenue, Longmont, CO 80501 (US).

(74) Agent: **YOERG, Darla, G.**; Sheridan Ross P.C., 1560 Broadway, Suite 1200, Denver, CO 80202 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
15 January 2009

(54) Title: CHIMERIC PUFA POLYKETIDE SYNTHASE SYSTEMS AND USES THEREOF

(57) Abstract: Disclosed are chimeric polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) proteins and chimeric PUFA PKS systems, including chimeric PUFA PKS proteins and systems derived from *Schizochytrium and Thraustochytrium.* Disclosed are nucleic acids and proteins encoding such chimeric PUFA PKS proteins and systems, genetically modified organisms comprising such chimeric PUFA PKS proteins and systems, and methods of making and using such chimeric PUFA PKS proteins and systems.

WO 2008/144473 A3

CHIMERIC PUFA POLYKETIDE SYNTHASE SYSTEMS AND USES THEREOF

Related Applications

Each of the following patent applications is incorporated herein by reference in its entirety. U.S. Patent Application No. 11/689,438, filed March 21, 2007; U.S. Patent Application No. 10/965,017, filed October 13, 2004, now U.S. Patent No. 7,217,856; U.S. Patent Application Serial No. 10/810,352, filed March 26, 2004, now U.S. Patent No. 7,211,418; U.S. Provisional Application Serial No. 60/457,979, filed March 26, 2003; U.S. Patent Application Serial No. 10/124,800, filed April 16, 2002; U.S. Provisional Application Serial No. 60/284,066, filed April 16, 2001; U.S. Provisional Application Serial No. 60/298,796, filed June 15, 2001; U.S. Provisional Application Serial No. 60/323,269, filed September 18, 2001; U.S. Application Serial No. 09/231,899, filed January 14, 1999, now U.S. Patent No. 6,566,583; U.S. Application No. 11/668,333, filed January 29, 2007; U.S. Patent Application Serial No. 11/452,096, filed June 12, 2006; U.S. Provisional Application No. 60/784,616, filed March 21, 2006; U.S. Provisional Application No. 60/689,167, filed June 10, 2005; U.S. Patent Application No. 11/452,138, filed June 12, 2006; U.S. Provisional Application No. 60/784,616, filed March 21, 2006; U.S. Provisional Application No. 60/689,167, filed June 10, 2005; U.S. Application Serial No. 09/090,793, filed June 4, 1998, now U.S. Patent No. 6,140,486.

Field of the Invention

This invention relates to chimeric polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems, and particularly, to chimeric PUFA PKS systems from *Schizochytrium* and *Thraustochytrium*. More particularly, this invention relates to nucleic acids encoding such PUFA PKS systems, to these PUFA PKS systems, to genetically modified organisms comprising such PUFA PKS systems, and to methods of making and using such PUFA PKS systems disclosed herein.

Background of the Invention

Polyketide synthase (PKS) systems are generally known in the art as enzyme complexes related to fatty acid synthase (FAS) systems, but which are often highly modified to produce specialized products that typically show little resemblance to fatty acids. It has now been shown, however, that PKS-like systems, also referred to herein interchangeably as PUFA PKS systems, PUFA synthase systems, or PKS systems for the production of PUFAs, exist in marine bacteria and certain eukaryotic organisms that are

capable of synthesizing polyunsaturated fatty acids (PUFAs) from acetyl-CoA and malonyl-CoA. The PUFA PKS pathways for PUFA synthesis in *Shewanella* and another marine bacteria, *Vibrio marinus*, are described in detail in U.S. Patent No. 6,140,486. The PUFA PKS pathways for PUFA synthesis in the eukaryotic Thraustochytrid, *Schizochytrium*, is described in detail in U.S. Patent No. 6,566,583. The PUFA PKS pathways for PUFA synthesis in eukaryotes such as members of Thraustochytriales, including the additional description of a PUFA PKS system in *Schizochytrium* and the identification of a PUFA PKS system in *Thraustochytrium*, including details regarding uses of these systems, are described in detail in U.S. Patent Application Publication No. 20020194641, published December 19, 2002, and U.S. Patent Application Publication No. 20070089199, published April 19, 2007. U.S. Patent Application Publication No. 20040235127, published November 25, 2004, discloses the detailed structural description of a PUFA PKS system in *Thraustochytrium*, and further detail regarding the production of eicosapentaenoic acid (C20:5, ω-3) (EPA) and other PUFAs using such systems. U.S. Patent Application Publication No. 20050100995, published May 12, 2005, discloses the structural and functional description of PUFA PKS systems in *Shewanella olleyana* and *Shewanella japonica,* and uses of such systems. These applications also disclose the genetic modification of organisms, including microorganisms and plants, with the genes comprising the PUFA PKS pathway and the production of PUFAs by such organisms. Furthermore, PCT Patent Publication No. WO 05/097982 describes a PUFA PKS system in *Ulkenia*, and U.S. Patent Application Publication No. 20050014231 describes PUFA PKS genes and proteins from *Thraustochytrium aureum*. Each of the above-identified applications is incorporated by reference herein in its entirety.

Researchers have attempted to exploit polyketide synthase (PKS) systems that have been traditionally described in the literature as falling into one of three basic types, typically referred to as: Type I (modular or iterative), Type II, and Type III. For purposes of clarity, it is noted that the Type I modular PKS system has previously also been referred to as simply a "modular" PKS system, and the Type I iterative PKS system has previously also been referred to simply as a "Type I" PKS system. The Type II system is characterized by separable proteins, each of which carries out a distinct enzymatic reaction. The enzymes work in concert to produce the end product and each individual enzyme of the system typically participates several times in the production of the end product. This type of system operates in a manner analogous to the fatty acid synthase

(FAS) systems found in plants and bacteria. Type I iterative PKS systems are similar to the Type II system in that the enzymes are used in an iterative fashion to produce the end product. The Type I iterative differs from Type II in that enzymatic activities, instead of being associated with separable proteins, occur as domains of larger proteins. This system is analogous to the Type I FAS systems found in animals and fungi.

In contrast to the Type II systems, in Type I modular PKS systems, each enzyme domain is used only once in the production of the end product. The domains are found in very large proteins and the product of each reaction is passed on to another domain in the PKS protein. Additionally, in the PKS systems described above, if a carbon-carbon double bond is incorporated into the end product, it is usually in the *trans* configuration.

Type III systems have been more recently discovered and belong to the plant chalcone synthase family of condensing enzymes. Type III PKSs are distinct from type I and type II PKS systems and utilize free acyl-CoA substrates in iterative condensation reactions to usually produce a heterocyclic end product.

Polyunsaturated fatty acids (PUFAs) are considered to be useful for nutritional, pharmaceutical, industrial, and other purposes. The current supply of PUFAs from natural sources and from chemical synthesis is not sufficient for commercial needs. A major current source for PUFAs is from marine fish; however, fish stocks are declining, and this may not be a sustainable resource. Additionally, contamination, from both heavy metals and toxic organic molecules, is a serious issue with oil derived from marine fish. Vegetable oils derived from oil seed crops are relatively inexpensive and do not have the contamination issues associated with fish oils. However, the PUFAs found in commercially developed plant oils are typically limited to linoleic acid (eighteen carbons with 2 double bonds, in the delta 9 and 12 positions - 18:2 delta 9,12) and linolenic acid (18:3 delta 9,12,15). In the conventional pathway (*i.e.*, the "standard" pathway or "classical" pathway) for PUFA synthesis, medium chain-length saturated fatty acids (products of a fatty acid synthase (FAS) system) are modified by a series of elongation and desaturation reactions. The substrates for the elongation reaction are fatty acyl-CoA (the fatty acid chain to be elongated) and malonyl-CoA (the source of the 2 carbons added during each elongation reaction). The product of the elongase reaction is a fatty acyl-CoA that has two additional carbons in the linear chain. The desaturases create *cis* double bonds in the preexisting fatty acid chain by extraction of 2 hydrogens in an oxygen-dependant reaction. The substrates for the desaturases are either acyl-CoA (in some

animals) or the fatty acid that is esterified to the glycerol backbone of a phospholipid (e.g. phosphatidylcholine).

Therefore, because a number of separate desaturase and elongase enzymes are required for fatty acid synthesis from linoleic and linolenic acids to produce the more unsaturated and longer chain PUFAs, engineering plant host cells for the expression of PUFAs such as EPA and docosahexaenoic acid (DRA) may require expression of several separate enzymes to achieve synthesis. Additionally, for production of useable quantities of such PUFAs, additional engineering efforts may be required. Therefore, it is of interest to obtain genetic material involved in PUF A biosynthesis from species that naturally produce these fatty acids *(e.g.,* from a PUFA PKS system) and to express the isolated material alone or in combination in a heterologous system which can be manipulated to allow production of commercial quantities of PUFAs.

There have been many efforts to produce PUFAs III oil-seed crop plants by modification of the endogenously-produced fatty acids. Genetic modification of these plants with various individual genes for fatty acid elongases and desaturases has produced leaves or seeds containing measurable levels of PUFAs such as EPA, but also containing significant levels of mixed shorter-chain and less unsaturated PUFAs (Qi et al., *Nature Biotech.* 22:739 (2004); PCT Publication No. WO 04/071467; Abbadi et al., *Plant Cell* 16:1 (2004)); Napier and Sayanova, Proceedings of the Nutrition Society (2005),64:387-20 393; Robert et al., Functional Plant Biology (2005) 32:473-479; or U.S. Patent Application Publication 2004/0172682.

Improvement in both microbial and plant production of PUFAs is a highly desirable commercial goal. Therefore, there remains a need in the art for a method to efficiently and effectively produce quantities of lipids *(e.g.,* triacylglycerol (TAG) and phospholipid (PL)) enriched in desired PUFAs, particularly in commercially useful organisms such as microorganisms and oil-seed plants.

Summary of the Invention

According to a first aspect there is provided a chimeric PUFA PKS system, wherein an FabA-like β-hydroxyacyl-ACP dehydrase-2 (DH2) domain from a first PUFA PKS system is replaced with a DH2 domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system.

According to a second aspect there is provided a method of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system, comprising expressing the chimeric PUFA PKS system of the first aspect in an organism.

According to a third aspect there is provided a genetically modified microorganism or plant or part of the plant, comprising a chimeric PUFA PKS system of the first aspect.

According to a fourth aspect there is provided a method of increasing the production of PUFAs and of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system, comprising expressing a chimeric PUFA PKS system in an organism, wherein the Fab A- like β-hydroxyacyl-ACP dehydrase-2 (DH2) domain from a first PUFA PKS system is replaced with a DH2 domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system, and wherein the DH2 domain from the second PUFA PKS system is optimized for the codon usage of the organism from which the first PUFA PKS system is derived.

According to a fifth aspect there is provided an isolated nucleic acid molecule encoding a chimeric OrfC protein that is at least 95% identical to SEQ ID NO: 74 or SEQ ID NO: 75.

According to a sixth aspect there is provided a recombinant nucleic acid molecule comprising the nucleic acid molecule of the fifth aspect.

According to a seventh aspect there is provided a recombinant host cell that has been transfected with the nucleic acid molecule of the fifth aspect. The host cell may be a plant cell or a microorganism such as a *Schizochytrium*, a bacterium, or a yeast.

According to an eighth aspect there is provided a genetically modified plant or part thereof, comprising the plant cell of the seventh aspect.

One embodiment of the invention relates to a chimeric PUFA PKS system, wherein an FabA-like β-hydroxyacyl-ACP dehydrase (DR) domain from a first PUFA PKS system is replaced with a DR domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system. In one aspect, a protein comprising the DR domain from the first PUFA PKS system is replaced with a

homologous protein comprising the DH domain from the second PUFA PKS system. In one aspect, the DH domain from the first or second PUFA PKS system corresponds to a DH2 domain from *Schizochytrium* or *Thraustochytrium*. In one aspect, the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein the second PUFA PKS system is a *Thraustochytrium* PUFA PKS system. In one aspect, the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein OrfC from the *Schizochytrium* PUFA PKS system is replaced with OrfC from a different thraustochytrid.

In one aspect of this embodiment, the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein OrfC from the *Schizochytrium* PUFA PKS system is replaced with OrfC from *Thraustochytrium* 23B. In one aspect, such an OrfC from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage. An exemplary nucleic acid sequence comprises SEQ ID NO:70. In an additional aspect, OrfA from the *Schizochytrium* PUFA PKS system is replaced with OrfA from *Thraustochytrium* 23B. In one aspect, such an OrfA from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage. An exemplary nucleic acid sequence comprises SEQ ID NO:71. In another additional aspect, OrfB from the *Schizochytrium* PUFA PKS system is replaced with OrfB from *Thraustochytrium* 23B. In one aspect, such an OrfB from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage. An exemplary nucleic acid sequence comprises SEQ ID NO:72. Other combinations of OrfsA, B and C will be apparent based on this disclosure to those of skill in the art.

In yet another aspect of this embodiment, the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and the DH2 domain of OrfC from the *Schizochytrium* PUFA PKS system is replaced with the DH2 domain from *Thraustochytrium* 23B. In one aspect, an exemplary nucleic acid sequence comprising the DH2 domain from *Thraustochytrium* 23B comprises SEQ ID NO:73. In one aspect, the DH2 domain from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage. Such a nucleic acid sequence comprising the DH2 domain from *Thraustochytrium* 23B is exemplified by the nucleic acid sequence comprising SEQ ID NO:75.

In yet another aspect of this embodiment, the chimeric PUFA PKS system comprises a protein comprising an amino acid sequence that is at least 95% identical to

SEQ ID NO:74. In one aspect, the chimeric PUFA PKS system comprises a protein comprising an amino acid sequence of SEQ ID NO:74. In one aspect, the chimeric PUFA PKS system comprises SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:74. In another aspect, the chimeric PUFA PKS system comprises SEQ ID NO:39, SEQ ID NO:4 and SEQ ID NO:62. In another aspect, the chimeric PUFA PKS system comprises SEQ ID NO:39, SEQ ID NO:4 and SEQ ID NO:74. In another aspect, the chimeric PUFA PKS system is encoded by nucleic acid molecules comprising: SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:70. In yet another aspect, the chimeric PUFA PKS system is encoded by nucleic acid molecules comprising: SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:73. In another aspect, the chimeric PUFA PKS system is encoded by nucleic acid molecules comprising: SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:75. In another aspect, the chimeric PUFA PKS system is encoded by nucleic acid molecules comprising: SEQ ID NO:71, SEQ ID NO:3 and SEQ ID NO:70.

Another embodiment of the invention relates to a method of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system, comprising expressing any of the above-described chimeric PUFA PKS systems in an organism. In one aspect, the chimeric PUFA PKS system is expressed by a microorganism. In one aspect, the microorganism is a *Schizochytrium*. In yet another aspect, the microorganism is a yeast. In one aspect, the chimeric PUFA PKS system is expressed by a plant.

Yet another embodiment of the invention relates to a genetically modified microorganism or plant or part of the plant, comprising any of the above-described chimeric PUFA PKS systems.

Another embodiment of the invention relates to a method of increasing the production of PUFAs and of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system. The method comprises expressing a chimeric PUFA PKS system in an organism, wherein the FabA-like β-hydroxyacyl-ACP dehydrase (DH) domain from a first PUFA PKS system is replaced with a DH domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system. The DH domain from the second PUFA PKS system is optimized for the codon usage of the organism from which the first PUFA PKS system is derived.

Yet another embodiment of the invention relates to an isolated nucleic acid molecule encoding a chimeric OrfC protein that is at least 95% identical to SEQ ID NO:74. In one aspect, the isolated nucleic acid molecule comprises a nucleic acid sequence that is at least 95% identical to SEQ ID NO:73. In one aspect, the nucleic acid sequence is optimized for the codon usage of an organism in which the nucleic acid molecule is to be expressed. As an example, the nucleic acid sequence may be optimized for the codon usage of an organism from which a portion of the chimeric protein is derived. In one embodiment, the nucleic acid sequence is at least 95% identical to SEQ ID NO:75.

Another embodiment of the invention relates to a recombinant nucleic acid molecule comprising any of the above-described nucleic acid molecules.

Yet another embodiment of the invention relates to a recombinant host cell that has been transfected with any of the above-described nucleic acid molecules. In one aspect, cell is a microorganism. In one aspect, the microorganism is a *Schizochytrium*. In one aspect, the microorganism is a bacterium. In one aspect, the microorganism is a yeast. In one aspect, the cell is a plant cell.

Another embodiment of the invention relates to a genetically modified plant or part thereof, comprising any of the above-described recombinant host cells.

Another embodiment of the invention relates to a chimeric PUFA PKS system, comprising: (a) at least one enoyl-ACP reductase (ER) domain; (b) at least four ACP domains; (c) at least two β-ketoacyl-ACP synthase (KS) domains; (d) at least one acyltransferase (AT) domain; (e) at least one β-ketoacyl-ACP reductase (KR) domain; (f) at least two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) at least one chain length factor (CLF) domain; and (h) at least one malonyl-CoA:ACP acyltransferase (MAT) domain. At least one of the DH domains is from a first PUFA PKS system, and the remainder of domains (a)-(h) are from a second, different PUFA PKS system.

Another embodiment of the invention relates to a method of increasing PUFA production by an organism that expresses a PUFA PKS system. The method includes modifying a nucleic acid molecule encoding at least one protein in the PUFA PKS system for the optimized codon usage of the organism or of a related organism. In one aspect, the organism expresses a heterologous, recombinant PUFA PKS system. In one aspect, the organism is a *Schizochytrium* and a nucleic acid molecule encoding at least one protein in the endogenous PUFA PKS system is optimized for *Schizochytrium* codon usage.

Brief Description of the Figures

Fig. 1 is a graphical representation of the domain structure of the *Schizochytrium* PUFA PKS system.

Fig. 2A is a schematic drawing showing step 1 of the construction of a plasmid containing a synthetic, *Schizochytrium* codon-optimized nucleic acid sequence encoding OrfC from *Thraustochytrium* 23B (pThOrfC_synPS), as well as intermediate plasmids produced by the process.

Fig. 2B is a schematic drawing showing step 2 of the construction of a plasmid containing a synthetic, *Schizochytrium* codon-optimized nucleic acid sequence encoding OrfC from *Thraustochytrium* 23B (pThOrfC_synPS), as well as intermediate plasmids produced by the process.

Fig. 3A is a schematic drawing showing steps 1-6 of the construction of a plasmid encoding *Schizochytrium* OrfC comprising a native DH2 domain from *Thraustochytrium* 23B (pDS49), as well as intermediate plasmids produced by the process.

Fig. 3B is a schematic drawing showing step 7 of the construction of a plasmid encoding *Schizochytrium* OrfC comprising a native DH2 domain from *Thraustochytrium* 23B (pDS49), as well as intermediate plasmids produced by the process.

Fig. 3C is a schematic drawing showing steps 8-9 of the construction of a plasmid encoding *Schizochytrium* OrfC comprising a native DH2 domain from *Thraustochytrium* 23B (pDS49), as well as intermediate plasmids produced by the process.

Fig. 4A is a schematic drawing showing the construction of plasmid DD21 as the first step in the construction of a plasmid encoding *Schizochytrium* OrfC comprising a synthetic, *Schizochytrium* codon-optimized DH2 domain from *Thraustochytrium* 23B (pDD24), as well as intermediate plasmids produced by the process.

Fig. 4B is a schematic drawing showing the construction of plasmid DD22 as the second step in the construction of a plasmid encoding *Schizochytrium* OrfC comprising a synthetic, *Schizochytrium* codon-optimized DH2 domain from *Thraustochytrium* 23B (pDD24), as well as intermediate plasmids produced by the process.

Fig. 4C is a schematic drawing showing the construction of plasmid pDD24 as the final step in the construction of a plasmid encoding *Schizochytrium* OrfC comprising a synthetic, *Schizochytrium* codon-optimized DH2 domain from *Thraustochytrium* 23B (pDD24), as well as intermediate plasmids produced by the process.

Fig. 5 is a FAME profile of control yeast and yeast expressing *Schizochytrium* OrfsA, OrfsB, OrfC and Het I.

Fig. 6 is the FAME profile for yeast from Fig. 5, expanded to illustrate the production of target PUFAs.

Detailed Description of the Invention

The present invention generally relates to polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems, also known as PUFA synthase systems, including PUFA PKS systems from thraustochytrids (*e.g., Schizochytrium* and *Thraustochytrium*), labyrinthulids, marine bacteria, and other PUFA PKS-containing organisms, and chimeric PUFA PKS proteins and systems produced therefrom. The present invention relates to genetically modified organisms comprising such PUFA PKS systems, and to methods of making and using such systems for the production of products of interest, including bioactive molecules. In one preferred embodiment, the present invention relates to a method to produce PUFAs in a microorganism or in an oil-seed plant or plant part that has been genetically modified to express a PUFA PKS system of the present invention. The oils produced by the microorganism or plant contain at least one PUFA produced by the PUFA PKS system, and in the case of the plant, are substantially free of the mixed shorter-chain and less unsaturated PUFAs that are fatty acid products produced by the modification of products of the FAS system. The present invention specifically includes methods to modify the amount of PUFAs and the ratio of PUFAs produced by a PUFA PKS system, and in one aspect of the invention, the ratio of omega-3 to omega-6 PUFAs or the ratio of one PUFA to another PUFA(s) (*e.g.,* the ratio of DHA to EPA), which can be applied to the creation and use of any PUFA PKS construct and/or genetically modified organism, as exemplified and described in detail herein.

First, the present inventors describe herein a domain of a PUFA PKS system that is both necessary and sufficient for modifying the ratio of PUFAs that are produced by a PUFA PKS system when more than one PUFA is produced, and provide novel chimeric constructs, novel chimeric PUFA PKS systems, novel organisms, and novel methods for producing modified amounts of PUFAs using this discovery. Second, the present inventors describe herein methods, modifications, and a variety of chimeric PUFA PKS systems and constructs for optimizing PUFA PKS expression in heterologous hosts (or in an endogenous host) to increase the PUFA production by the organism. The invention

includes a detailed description of the use of these two discoveries, alone or together, to enhance and direct PUFA production in an organism.

More particularly, with regard to certain embodiments of the invention, previous work by the present inventors and colleagues (see Example 8 in U.S. Patent Application Publication No. 20050100995) demonstrated that the *Thraustochytrium* 23B orfC coding region (represented herein by SEQ ID NO:62) could functionally replace the *Schizochytrium* orfC coding region in the orfC locus in the genome. This was determined by first creating an exact deletion of the *Schizochytrium* orfC coding region containing an antibiotic resistance cassette in its place (denoted ΔorfC::ZEO) resulting in a strain (denoted B32-Z1) with an obligate growth requirement for DHA and resistance to Zeocin™. A plasmid in which the *Th.*23B orfC coding region was cloned exactly between *Schizochytrium* orfC upstream and downstream non-coding regions was then constructed. Transformation of the *Schizochytrium* ΔorfC::ZEO strain with this *Th.*23B orfC construct resulted in complementation of the deletion and prototrophic (non-DHA-requiring), Zeocin-sensitive transformants. It was determined that these transformants derived from double cross-over recombination events at the orfC locus such that the *Th.*23B orfC coding region had exactly substituted for that from *Schizochytrium*; *i.e.,* gene replacement. Analysis of the fatty acid content of these transformants showed that the DHA/DPA ratio had been changed from *ca.* 2.3 (in wild type *Schizochytrium* ATCC20888) to *ca.* 8.3 (approximately that of *Th.*23B). This result indicated that the orfC gene (containing three domains, DH1, DH2 and ER, in *Schizochytrium* and *Thraustochytrium*) plays a major role in determining the n-3/n-6 (omega-3/omega-6) ratio of PUFA products. However, total PUFA production in the *Th.*23B orfC-containing strain, while significant, was lower than that of the wild-type *Schizochytrium* host (*ca.* 60%).

Examination of these two orfC coding regions led the inventors to consider that the *Th.*23B gene is poorly expressed in *Schizochytrium* due to notably different patterns of codon usage between *Schizochytrium* and *Thraustochytrium*. The inventors have now discovered that by using a "synthetic" *Th.*23B orfC coding region (*i.e.,* a synthetically produced coding region) with codon usage optimized for the *Schizochytrium* pattern, DHA production was enhanced, while the increased n-3/n-6 ratio seen with the non-synthetic *Th.*23B orfC was maintained (see Examples 1 and 4).

The inventors have also previously described the existence of identifiable domains within the OrfC protein for *Schizochytrium* and *Thraustochytrium*: dehydratase 1 (DH1),

dehydratase 2 (DH2), and enoyl reductase (ER) (*e.g.,* see U.S. Patent Application Publication No. 20020194641, *supra*; U.S. Patent Application Publication No. 20040235127, *supra*), and have taught that one or more of the domains in OrfC were believed to be involved in controlling the type and/or ratio of fatty acids produced by the PUFA PKS system. Here, the inventors demonstrate in *Schizochytrium*, *E. coli*, and yeast systems that the DH2 domain *alone* is responsible for most or all of the effect of the PUFA PKS system on the omega-3 to omega-6 (n-3/n-6) fatty acid ratio. In particular, the inventors first performed experiments in which various *Thraustochytrium* 23B OrfC domains were used to replace the corresponding domains in *Schizochytrium* OrfC (data not shown). The inventors found that replacement of the *Schizochytrium* OrfC-ER domain with that from *Thraustochytrium* did not significantly change the DHA/DPA ratio as compared to wild-type *Schizochytrium* (historically, approximately 2.3). However, replacement of both *Schizochytrium* DH domains with the corresponding domains from *Thraustochytrium* significantly increased the DHA/DPA ratio toward that of wild-type *Thraustochytrium* 23B (historically, approximately 8.3-10), and replacement of just the *Schizochytrium* DH2 domain with that from *Thraustochytrium* 23B, was sufficient to achieve effectively the same result. Examples 2, 3, 4, 5, and 6 provide a variety of experimental results demonstrating the effect of the DH2 domain on the omega-3 to omega-6 (n-3/n-6) fatty acid ratio in PUFA PKS systems.

The present inventors also describe the use of a variety of chimeric PUFA PKS systems to increase the production of PUFAs by the host organism, and have made the unexpected discovery that certain chimeric PUFA PKS combinations (*e.g.,* chimeric PUFA PKS systems comprised of particular combinations of Orfs from *Schizochytrium* and *Thraustochytrium*) have significantly higher PUFA production, and in one example, DHA production, than the native organisms or than other chimeric PUFA PKS systems. For example, the inventors demonstrate that a chimeric PUFA PKS system comprised of an OrfA and OrfC from *Thraustochytrium* 23B and an OrfB from *Schizochytrium*, when expressed in a *Schizochytrium* host organism, produces significantly more fatty acids and significantly more DHA specifically, than native *Schizochytrium* or than other chimeric PUFA PKS systems derived from these two organisms (Example 8). Accordingly, the invention provides substantial guidance on the production of several different PUFA PKS systems that have increased PUFA production and improved n-3/n-6 ratios, as compared to some wild-type (non-chimeric) PUFA synthases.

As used herein, a PUFA PKS system (which may also be referred to as a PUFA synthase system, a PUFA synthase, or a PKS-like system for the production of PUFAs) generally has the following identifying features: (1) it produces PUFAs, and particularly, long chain PUFAs, as a natural product of the system; and (2) it comprises several multifunctional proteins assembled into a complex that conducts both *iterative* processing of the fatty acid chain as well *non-iterative* processing, including *trans-cis* isomerization and enoyl reduction reactions in selected cycles. In addition, the ACP domains present in the PUFA synthase enzymes require activation by attachment of a cofactor (4-phosphopantetheine). Attachment of this cofactor is carried out by phosphopantetheinyl transferases (PPTase). If the endogenous PPTases of the host organism are incapable of activating the PUFA synthase ACP domains, then it is necessary to provide a PPTase that is capable of carrying out that function. The inventors have identified the Het I enzyme of *Nostoc* sp. as an exemplary and suitable PPTase for activating PUFA synthase ACP domains. Reference to a PUFA PKS system or a PUFA synthase refers collectively to all of the genes and their encoded products that work in a complex to produce PUFAs in an organism. Therefore, the PUFA PKS system refers specifically to a PKS system for which the natural products are PUFAs.

More specifically, a PUFA PKS system as referenced herein produces polyunsaturated fatty acids (PUFAs) and particularly, long chain PUFAs, as products. For example, an organism that endogenously (naturally) contains a PUFA PKS system makes PUFAs using this system. According to the present invention, PUFAs are fatty acids with a carbon chain length of at least 16 carbons, and more preferably at least 18 carbons, and more preferably at least 20 carbons, and more preferably 22 or more carbons, with at least 3 or more double bonds, and preferably 4 or more, and more preferably 5 or more, and even more preferably 6 or more double bonds, wherein all double bonds are in the *cis* configuration. Reference to long chain polyunsaturated fatty acids (LCPUFAs) herein more particularly refers to fatty acids of 18 and more carbon chain length, and preferably 20 and more carbon chain length, containing 3 or more double bonds. LCPUFAs of the omega-6 series include: gamma-linolenic acid (C18:3), di-homo-gamma-linolenic acid (C20:3n-6), arachidonic acid (C20:4n-6), adrenic acid (also called docosatetraenoic acid or DTA) (C22:4n-6), and docosapentaenoic acid (C22:5n-6). The LCPUFAs of the omega-3 series include: alpha-linolenic acid (C18:3), eicosatrienoic acid (C20:3n-3), eicosatetraenoic acid (C20:4n-3), eicosapentaenoic acid (C20:5n-3), docosapentaenoic

acid (C22:5n-3), and docosahexaenoic acid (C22:6n-3). The LCPUFAs also include fatty acids with greater than 22 carbons and 4 or more double bonds including but not limited to C28:8(n-3).

Second, a PUFA PKS system according to the present invention comprises several multifunctional proteins (and can include single function proteins, particularly for PUFA PKS systems from marine bacteria) that are assembled into a complex that conducts both *iterative* processing of the fatty acid chain as well *non-iterative* processing, including *trans-cis* isomerization and enoyl reduction reactions in selected cycles. These proteins can also be referred to herein as the core PUFA PKS enzyme complex or the core PUFA PKS system. The general functions of the domains and motifs contained within these proteins are individually known in the art and have been described in detail with regard to various PUFA PKS systems from marine bacteria and eukaryotic organisms (see, *e.g.,* U.S. Patent No. 6,140,486; U.S. Patent No. 6,566,583; Metz et al., *Science* 293:290-293 (2001); U.S. Patent Application Publication No. 20020194641; U.S. Patent Application Publication No. 20040235127; U.S. Patent Application Publication No. 20050100995, and PCT Publication No. WO 2006/135866). The domains may be found as a single protein (i.e., the domain and protein are synonymous) or as one of two or more (multiple) domains in a single protein, as mentioned above.

Before the discovery of a PUFA PKS system in marine bacteria (see U.S. Patent No. 6,140,486), PKS systems were not known to possess this combination of iterative and selective enzymatic reactions, and they were not thought of as being able to produce carbon-carbon double bonds in the *cis* configuration. However, the PUFA PKS system described by the present invention has the capacity to introduce *cis* double bonds and the capacity to vary the reaction sequence in the cycle.

The present inventors propose to use these features of the PUFA PKS system to produce a range of bioactive molecules that could not be produced by the previously described (Type I iterative or modular, Type II, or Type III) PKS systems. These bioactive molecules include, but are not limited to, polyunsaturated fatty acids (PUFAs), antibiotics or other bioactive compounds, many of which will be discussed below. For example, using the knowledge of the PUFA PKS gene structures described herein, any of a number of methods can be used to alter the PUFA PKS genes, or combine portions of these genes with other synthesis systems, including other PKS systems, such that new products are produced. The inherent ability of this particular type of system to do both

iterative and selective reactions will enable this system to yield products that would not be found if similar methods were applied to other types of PKS systems.

Preferably, a PUFA PKS system of the present invention comprises at least the following biologically active domains that are typically contained on three or more proteins: (a) at least one enoyl-ACP reductase (ER) domain; (b) multiple acyl carrier protein (ACP) domain(s) (*e.g.,* at least from one to four, and preferably at least five ACP domains, and in some embodiments up to six, seven, eight, nine, ten, or more than ten ACP domains); (c) at least two β-ketoacyl-ACP synthase (KS) domains; (d) at least one acyltransferase (AT) domain; (e) at least one β-ketoacyl-ACP reductase (KR) domain; (f) at least two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) at least one chain length factor (CLF) domain; (h) at least one malonyl-CoA:ACP acyltransferase (MAT) domain. In one embodiment, a PUFA PKS system according to the present invention also comprises at least one region containing a dehydratase (DH) conserved active site motif.

In one embodiment, a *Schizochytrium* PUFA PKS system comprises at least the following biologically active domains: (a) two enoyl-ACP reductase (ER) domain; (b) between four or five and ten or more acyl carrier protein (ACP) domains, and in one aspect, nine ACP domains; (c) two β-ketoacyl-ACP synthase (KS) domains; (d) one acyltransferase (AT) domain; (e) one β-ketoacyl-ACP reductase (KR) domain; (f) two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) one chain length factor (CLF) domain; and (h) one malonyl-CoA:ACP acyltransferase (MAT) domain. In one embodiment, a *Schizochytrium* PUFA PKS system according to the present invention also comprises at least one region or domain containing a dehydratase (DH) conserved active site motif that is not a part of a FabA-like DH domain. The structural and functional characteristics of these domains are generally individually known in the art and will be described in detail below with regard to the PUFA PKS systems of the present invention.

In another preferred embodiment, a *Thraustochytrium* PUFA PKS system comprises at least the following biologically active domains: (a) two enoyl-ACP reductase (ER) domain; (b) between four or five and ten or more acyl carrier protein (ACP) domains, and in one aspect, eight ACP domains; (c) two β-ketoacyl-ACP synthase (KS) domains; (d) one acyltransferase (AT) domain; (e) one β-ketoacyl-ACP reductase (KR) domain; (f) two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) one chain length factor (CLF) domain; and (h) one malonyl-CoA:ACP acyltransferase (MAT) domain. In one embodiment, a *Thraustochytrium* PUFA PKS system according to the

present invention also comprises at least one region or domain containing a dehydratase (DH) conserved active site motif that is not a part of a FabA-like DH domain. The structural and functional characteristics of these domains are generally individually known in the art and will be described in detail below with regard to the PUFA PKS systems of the present invention.

A PUFA PKS system can additionally include one or more accessory proteins, which are defined herein as proteins that are not considered to be part of the core PUFA PKS system as described above (*i.e.,* not part of the PUFA synthase enzyme complex itself), but which may be, or are, necessary for PUFA production or at least for efficient PUFA production using the core PUFA synthase enzyme complex of the present invention, particularly in certain host organisms (*e.g.,* plants). For example, in order to produce PUFAs, a PUFA PKS system must work with an accessory protein that transfers a 4'-phosphopantetheinyl moiety from coenzyme A to the acyl carrier protein (ACP) domain(s). Therefore, a PUFA PKS system can be considered to include at least one 4'-phosphopantetheinyl transferase (PPTase) domain, or such a domain can be considered to be an accessory domain or protein to the PUFA PKS system. When genetically modifying organisms (*e.g.,* microorganisms or plants) to express a PUFA PKS system according to the present invention, some host organisms may endogenously express accessory proteins that are needed to work with the PUFA PKS to produce PUFAs (*e.g.,* PPTases). However, some organisms may be transformed with nucleic acid molecules encoding one or more accessory proteins described herein to enable and/or to enhance production of PUFAs by the organism, even if the organism endogenously produces a homologous accessory protein (*i.e.,* some heterologous accessory proteins may operate more effectively or efficiently with the transformed PUFA synthase proteins than the host cells' endogenous accessory protein). The present invention and prior applications provide examples of bacteria and yeast that have been genetically modified with the PUFA PKS system of the present invention that includes an accessory PPTase. Plants that have been genetically modified with the PUFA PKS system that includes an accessory PPTase have been described (see, *e.g.,* U.S. Patent Application Publication No. 20070089199). Structural and functional characteristics of PPTases will be described in more detail below.

The "standard" or "classical" pathway for synthesis of long chain PUFAs (LCPUFAs) in eukaryotic organisms involves the modification of medium chain-length saturated or mono-unsaturated fatty acids (*e.g.,* the products of the FAS system described

above). These modifications consist of elongation steps and desaturation steps. The substrates for the elongation reaction are fatty acyl-CoA (the fatty acid chain to be elongated) and malonyl-CoA (the source of the two carbons added during each elongation reaction). The product of the elongase reaction is a fatty acyl-CoA that has two additional carbons in the linear chain. Free fatty acids (FFAs) do not normally occur in this reaction cycle. The desaturases create *cis* double bonds in the preexisting fatty acid chain by extraction of two hydrogens in an oxygen-dependant reaction. The substrates for the desaturases are either acyl-CoAs (in some animals) or fatty acids that are esterified to the glycerol backbone of a PL (e.g., phosphotidylcholine). Again, FFAs do not occur in this reaction mechanism. Therefore, the only time FFAs occur in "standard" or "classical" LCPUFA synthesis pathways is during release of the fatty acids from some FAS systems. As discussed above, these are typically 16 or 18 carbon fatty acids and usually are either saturated or monounsaturated fatty acids, not longer chain PUFAs such as EPA or DHA. One consequence of this scheme for long chain PUFA production is that intermediates in the pathway often accumulate, often representing the majority of the novel fatty acids produced by the system.

Therefore, according to the present invention, reference to a "standard" or "classical" pathway for the production of PUFAs refers to the fatty acid synthesis pathway where medium chain-length saturated fatty acids (*e.g.,* products of a fatty acid synthase (FAS) system) are modified by a series of elongation and desaturation reactions. The substrates for the elongation reaction are fatty acyl-CoA (the fatty acid chain to be elongated) and malonyl-CoA (the source of the 2 carbons added during each elongation reaction). The product of the elongase reaction is a fatty acyl-CoA that has two additional carbons in the linear chain. The desaturases create *cis* double bonds in the preexisting fatty acid chain by extraction of 2 hydrogens in an oxygen-dependant reaction. Such pathways and the genes involved in such pathways are well-known in the literature.

As used herein, the term “lipid” includes phospholipids (PL); free fatty acids; esters of fatty acids; triacylglycerols (TAG); diacylglycerides; monoacylglycerides; phosphatides; waxes (esters of alcohols and fatty acids); sterols and sterol esters; carotenoids; xanthophylls (*e.g.*, oxycarotenoids); hydrocarbons; and other lipids known to one of ordinary skill in the art. The terms "polyunsaturated fatty acid" and "PUFA" include not only the free fatty acid form, but other forms as well, such as the TAG form and the PL form.

Reference to a "heterologous" organism or "heterologous" host, with respect to the expression of a PUFA PKS protein, domain or system by the organism/host, means that at least one protein, domain, or portion of the PUFA PKS system is not a protein, domain or portion that is naturally (endogenously) expressed by the organism, although the PUFA PKS system may include proteins, domains, or portions thereof that are naturally expressed by host organism (*e.g.,* a chimeric protein as described herein that contains sequences derived from the host organism and from a different organism or different protein).

Certain exemplary nucleic acid molecules (constructs) encoding various chimeric proteins are described herein (see Examples). According to the present invention, a "chimeric protein" is an engineered protein encoded by a nucleic acid sequence that is produced by splicing or linking (ligating) together two or more complete or partial genes or nucleic acid sequences. A "chimeric PUFA PKS system" is a PUFA PKS system that contains proteins and/or domains, including chimeric proteins and/or domains, from two or more different PKS systems. For example, the Examples describe a chimeric PUFA PKS system comprised of the *Schizochytrium* PUFA PKS OrfA and OrfB and the *Thraustochytrium* PUFA PKS OrfC. The Examples also describe a chimeric PUFA PKS system comprised of the *Schizochytrium* PUFA PKS OrfA, OrfB, and all of OrfC except for the DH2 domain, which is the PUFA PKS DH2 domain from a *Thraustochytrium* PUFA PKS. This latter chimeric PUFA PKS system accordingly comprises a chimeric protein (a chimeric OrfC protein). The same chimeras are also described using *Thraustochytrium* nucleic acid sequences that have been optimized for *Schizochytrium* codon usage, illustrating a combination of genetic manipulations that can be used to alter the product produced by a PUFA PKS system (see Examples). The Examples also describe a variety of other chimeric PUFA PKS systems.

As used herein, "codon optimization" or derivative phrases thereof refer to the process of modifying (altering, changing, mutating) a nucleic acid sequence encoding a given protein to replace one or more codons in the sequence with codons that are most frequently used in nucleic acid sequences of a particular organism in which a nucleic acid molecule comprising the nucleic acid sequence is to be expressed. Codon bias and the general idea of codon optimization are understood by the skilled artisan. More particularly, the degree to which a given codon appears in the genetic code can vary significantly between organisms (*e.g.,* including from species to species within a genus).

Any codon that an organism uses a small percentage of the time, or less than another codon for the same amino acid, can cause problems with protein expression. Accordingly, protein expression can improve dramatically when the codon frequency of the nucleic acid sequence being used is matched to that of the host expression system/organism (*e.g.,* by replacing rare or infrequent or less frequently used codons with others that more closely reflect the host system's natural codon bias, without modifying the amino acid sequence).

The present inventors describe herein methods to optimize codon usage of a nucleic acid sequence for that of *Schizochytrium*, although this is just one example of the use of codon optimization in the present invention. According to the present invention, the nucleotide sequence of a nucleic acid molecule encoding a given protein (*e.g.,* a PUFA PKS protein) can be modified (*e.g.,* by synthesis, mutation, recombinant technology, etc.) for the optimal (optimized) codon usage of a host cell or organism in which the nucleic acid molecule is to be expressed, or indeed, for the optimized codon usage of a different organism (*e.g.,* a nucleic acid molecule encoding a *Thraustochytrium* PUFA PKS protein for expression in a *plant* may be optimized for *Schizochytrium* codon usage). Table 1 of the Examples illustrates optimized codon usage for *Schizochytrium*.

In addition, the inventors propose herein the optimization of the nucleic acid sequence of a nucleic acid molecule encoding a given protein for the *same host* from which the nucleic acid sequence was derived, learned or obtained, for expression *in that host* (or in another host). This latter embodiment of the invention represents a "directed" or "accelerated" evolution of sorts, in which, for example, a nucleic acid molecule encoding a protein from an organism (*e.g.,* a PUFA PKS protein from *Schizochytrium*) is modified (*e.g.,* by resynthesizing the nucleic acid sequence and replacing certain nucleotides) to enhance codon usage (optimize the codon usage) that is preferred by the same organism (*Schizochytrium*, in this example). This nucleic acid molecule can then be expressed in *Schizochytrium* (as a recombinant nucleic acid molecule) or in another host cell or organism (*e.g.,* in a plant). In this embodiment, it is proposed that a given nucleic acid sequence from an organism may not use the optimal codons (codon bias) that can be determined for that organism. Accordingly, one may resynthesize the nucleic acid sequence to improve protein expression in that organism.

PUFA PKS systems and proteins or domains thereof that are useful in the present invention include both bacterial and non-bacterial PUFA PKS systems. A non-bacterial PUFA PKS system is a PUFA PKS system that is from or derived from an organism that is

not a bacterium, such as a eukaryote or an archaebacterium. Eukaryotes are separated from prokaryotes based on the degree of differentiation of the cells, with eukaryotes being more differentiated than prokaryotes. In general, prokaryotes do not possess a nuclear membrane, do not exhibit mitosis during cell division, have only one chromosome, contain 70S ribosomes in their cytoplasm, do not possess mitochondria, endoplasmic reticulum, chloroplasts, lysosomes or Golgi apparatus, and may have flagella, which if present, contain a single fibril. In contrast, eukaryotes have a nuclear membrane, exhibit mitosis during cell division, have many chromosomes, contain 80S ribosomes in their cytoplasm, possess mitochondria, endoplasmic reticulum, chloroplasts (in algae), lysosomes and Golgi apparatus, and may have flagella, which if present, contain many fibrils. In general, bacteria are prokaryotes, while algae, fungi, protist, protozoa and higher plants are eukaryotes. According to the present invention, genetically modified organisms can be produced which incorporate non-bacterial PUFA PKS functional domains with bacterial PUFA PKS functional domains, as well as PKS functional domains or proteins from other PKS systems (Type I iterative or modular, Type II, or Type III) or FAS systems.

According to the present invention, a domain or protein having 3-keto acyl-ACP synthase (KS) biological activity (function) is characterized as the enzyme that carries out the initial step of the FAS (and PKS) elongation reaction cycle. The term "β-ketoacyl-ACP synthase" can be used interchangeably with the terms "3-keto acyl-ACP synthase", "β-keto acyl-ACP synthase", and "keto-acyl ACP synthase", and similar derivatives. The acyl group destined for elongation is linked to a cysteine residue at the active site of the enzyme by a thioester bond. In the multi-step reaction, the acyl-enzyme undergoes condensation with malonyl-ACP to form -keto acyl-ACP, CO_2 and free enzyme. The KS plays a key role in the elongation cycle and in many systems has been shown to possess greater substrate specificity than other enzymes of the reaction cycle. For example, *E. coli* has three distinct KS enzymes - each with its own particular role in the physiology of the organism (Magnuson et al., *Microbiol. Rev.* **57**, 522 (1993)). The two KS domains of the PUFA-PKS systems described in marine bacteria and the thraustochytrids described herein may have distinct roles in the PUFA biosynthetic reaction sequence. As a class of enzymes, KS's have been well characterized. The sequences of many verified KS genes are known, the active site motifs have been identified and the crystal structures of several have been determined. Proteins (or domains of proteins) can be readily identified as belonging to the KS family of enzymes by homology to known KS sequences.

According to the present invention, a domain or protein having malonyl-CoA:ACP acyltransferase (MAT) biological activity (function) is characterized as one that transfers the malonyl moiety from malonyl-CoA to ACP. The term "malonyl-CoA:ACP acyltransferase" can be used interchangeably with "malonyl acyltransferase" and similar derivatives. In addition to the active site motif (GxSxG), these enzymes possess an extended motif of R and Q amino acids in key positions that identifies them as MAT enzymes (*e.g.,* in contrast to an AT domain described below). In some PKS systems (but not the PUFA PKS domain) MAT domains will preferentially load methyl- or ethyl-malonate on to the ACP group (from the corresponding CoA ester), thereby introducing branches into the linear carbon chain. MAT domains can be recognized by their homology to known MAT sequences and by their extended motif structure.

According to the present invention, a domain or protein having acyl carrier protein (ACP) biological activity (function) is characterized as being small polypeptides (typically, 80 to 100 amino acids long), that function as carriers for growing fatty acyl chains via a thioester linkage to a covalently bound co-factor of the protein. They occur as separate units or as domains within larger proteins. ACPs are converted from inactive apo-forms to functional holo-forms by transfer of the phosphopantetheinyl moiety of CoA to a highly conserved serine residue of the ACP. Acyl groups are attached to ACP by a thioester linkage at the free terminus of the phosphopantetheinyl moiety. ACPs can be identified by labeling with radioactive pantetheine and by sequence homology to known ACPs. The presence of variations of the above mentioned motif (LGIDS*) is also a signature of an ACP.

According to the present invention, a domain or protein having ketoreductase activity, also referred to as 3-ketoacyl-ACP reductase (KR) biological activity (function), is characterized as one that catalyzes the pyridine-nucleotide-dependent reduction of 3-keto acyl forms of ACP. It is the first reductive step in the *de novo* fatty acid biosynthesis elongation cycle and a reaction often performed in polyketide biosynthesis. The term "β-ketoacyl-ACP reductase" can be used interchangeably with the terms "ketoreductase", "3-ketoacyl-ACP reductase", "keto-acyl ACP reductase" and similar derivatives of the term. Significant sequence similarity is observed with one family of enoyl ACP reductases (ER), the other reductase of FAS (but not the ER family present in the PUFA PKS systems), and the short-chain alcohol dehydrogenase family. Pfam analysis of the PUFA PKS region indicated above reveals the homology to the short-chain alcohol dehydrogenase family in

the core region. Blast analysis of the same region reveals matches in the core area to known KR enzymes as well as an extended region of homology to domains from the other characterized PUFA PKS systems.

According to the present invention, a domain or protein is referred to as a chain length factor (CLF) based on the following rationale. The CLF was originally described as characteristic of Type II (dissociated enzymes) PKS systems and was hypothesized to play a role in determining the number of elongation cycles, and hence the chain length, of the end product. CLF amino acid sequences show homology to KS domains (and are thought to form heterodimers with a KS protein), but they lack the active site cysteine. CLF's role in PKS systems has been controversial. New evidence (C. Bisang et al., *Nature* **401**, 502 (1999)) suggests a role in priming (providing the initial acyl group to be elongated) the PKS systems. In this role the CLF domain is thought to decarboxylate malonate (as malonyl-ACP), thus forming an acetate group that can be transferred to the KS active site. This acetate therefore acts as the 'priming' molecule that can undergo the initial elongation (condensation) reaction. Homologues of the Type II CLF have been identified as 'loading' domains in some modular PKS systems. A domain with the sequence features of the CLF is found in all currently identified PUFA PKS systems and in each case is found as part of a multidomain protein.

An "acyltransferase" or "AT" refers to a general class of enzymes that can carry out a number of distinct acyl transfer reactions. The term "acyltransferase" can be used interchangeably with the term "acyl transferase". The AT domains identified in the PUFA PKS systems described herein show good homology one another and to domains present in all of the other PUFA PKS systems currently examined and very weak homology to some acyltransferases whose specific functions have been identified (e.g. to malonyl-CoA:ACP acyltransferase, MAT). In spite of the weak homology to MAT, this AT domain is not believed to function as a MAT because it does not possess an extended motif structure characteristic of such enzymes (see MAT domain description, above). For the purposes of this disclosure, the possible functions of the AT domain in a PUFA PKS system include, but are not limited to: transfer of the fatty acyl group from the ORFA ACP domain(s) to water (i.e. a thioesterase - releasing the fatty acyl group as a free fatty acid), transfer of a fatty acyl group to an acceptor such as CoA, transfer of the acyl group among the various ACP domains, or transfer of the fatty acyl group to a lipophilic acceptor molecule (e.g. to lysophosphadic acid).

According to the present invention, this domain has enoyl reductase (ER) biological activity. The ER enzyme reduces the *trans*-double bond (introduced by the DH activity) in the fatty acyl-ACP, resulting in fully saturating those carbons. The ER domain in the PUFA-PKS shows homology to a newly characterized family of ER enzymes (Heath et al., *Nature* **406**, 145 (2000)). Heath and Rock identified this new class of ER enzymes by cloning a gene of interest from *Streptococcus pneumoniae*, purifying a protein expressed from that gene, and showing that it had ER activity in an *in vitro* assay. All of the PUFA PKS systems currently examined contain at least one domain with very high sequence homology to the *Schizochytrium* ER domain, which shows homology to the *S. pneumoniae* ER protein.

According to the present invention, a protein or domain having dehydrase or dehydratase (DH) activity catalyzes a dehydration reaction. As used generally herein, reference to DH activity typically refers to FabA-like β-hydroxyacyl-ACP dehydrase (DH) biological activity. FabA-like β-hydroxyacyl-ACP dehydrase (DH) biological activity removes HOH from a β-ketoacyl-ACP and initially produces a *trans* double bond in the carbon chain. The term "FabA-like β-hydroxyacyl-ACP dehydrase" can be used interchangeably with the terms "FabA-like β-hydroxy acyl-ACP dehydrase", "β-hydroxyacyl-ACP dehydrase", "dehydrase" and similar derivatives. The DH domains of the PUFA PKS systems show homology to bacterial DH enzymes associated with their FAS systems (rather than to the DH domains of other PKS systems). A subset of bacterial DH's, the FabA-like DH's, possesses *cis-trans* isomerase activity (Heath et al., *J. Biol. Chem.*, 271, 27795 (1996)). It is the homology to the FabA-like DH proteins that suggests that one or all of the DH domains described herein is responsible for insertion of the cis double bonds in the PUFA PKS products.

A PUFA PKS protein useful of the invention may also have dehydratase activity that is not characterized as FabA-like (e.g., the *cis-trans* activity described above is associated with FabA-like activity), generally referred to herein as non-FabA-like DH activity, or non-FabA-like β-hydroxyacyl-ACP dehydrase (DH) biological activity. More specifically, a conserved active site motif (~13 amino acids long: L*xxHxxxGxxxxP; *in the motif, L can also be I) is found in dehydratase domains in PKS systems (Donadio S, Katz L. Gene. 1992 Feb 1;111(1):51-60). This conserved motif, also referred to herein as a dehydratase (DH) conserved active site motif or DH motif, is found in a similar region of all known PUFA-PKS sequences described to date and in the PUFA PKS sequences

described herein, but it is believed that his motif has only recently been detected. This conserved motif is within an uncharacterized region of high homology in the PUFA-PKS sequence. The proposed biosynthesis of PUFAs via the PUFA-PKS requires a non-FabA like dehydration, and this motif may be associated with that reaction.

For purposes of illustration, the structure of certain PUFA PKS systems is described in detail below. However, it is to be understood that this invention is not limited to the use of these PUFA PKS systems. For example, a detailed description of bacterial PUFA PKS systems can be found in U.S. Patent No. 6,140,486 and U.S. Patent Application Publication No. 20050100995, and a description of other PUFA PKS genes or systems is found in PCT Patent Publication No. WO 05/097982 and U.S. Patent Application Publication No. 20050014231.

***Schizochytrium* PUFA PKS System**

Schizochytrium is a thraustochytrid marine microorganism that accumulates large quantities of triacylglycerols rich in DHA and docosapentaenoic acid (DPA; 22:5 ω-6); *e.g.*, 30% DHA + DPA by dry weight (Barclay et al., *J. Appl. Phycol.* **6**, 123 (1994)). In eukaryotes that synthesize 20- and 22-carbon PUFAs by an elongation/desaturation pathway, the pools of 18-, 20- and 22-carbon intermediates are relatively large so that *in vivo* labeling experiments using [^{14}C]-acetate reveal clear precursor-product kinetics for the predicted intermediates (Gellerman et al., *Biochim. Biophys. Acta* 573:23 (1979)). Furthermore, radiolabeled intermediates provided exogenously to such organisms are converted to the final PUFA products. The present inventors have shown that [1-^{14}C]-acetate was rapidly taken up by *Schizochytrium* cells and incorporated into fatty acids, but at the shortest labeling time (1 min), DHA contained 31% of the label recovered in fatty acids, and this percentage remained essentially unchanged during the 10-15 min of [^{14}C]-acetate incorporation and the subsequent 24 hours of culture growth (See U.S. Patent Application Publication No. 20020194641, *supra*). Similarly, DPA represented 10% of the label throughout the experiment. There is no evidence for a precursor-product relationship between 16- or 18-carbon fatty acids and the 22-carbon polyunsaturated fatty acids. These results are consistent with rapid synthesis of DHA from [^{14}C]-acetate involving very small (possibly enzyme-bound) pools of intermediates.

Fig. 1 is a graphical representation of the three open reading frames from the *Schizochytrium* PUFA PKS system, and includes the domain structure of this PUFA PKS system. There are three open reading frames that form the core *Schizochytrium* PUFA PKS system. The domain structure of each open reading frame is as follows.

Schizochytrium Open Reading Frame A (OrfA):

The complete nucleotide sequence for OrfA is represented herein as SEQ ID NO:1. OrfA is a 8730 nucleotide sequence (not including the stop codon) which encodes a 2910 amino acid sequence, represented herein as SEQ ID NO:2. Within OrfA are twelve domains: (a) one β-keto acyl-ACP synthase (KS) domain; (b) one malonyl-CoA:ACP acyltransferase (MAT) domain; (c) nine acyl carrier protein (ACP) domains; and (d) one ketoreductase (KR) domain. Genomic DNA clones (plasmids) encoding OrfA from both *Schizochytrium* sp. ATCC 20888 and a daughter strain of ATCC 20888, denoted *Schizochytrium* sp., strain N230D, have been isolated and sequenced. N230D was one of more than 1,000 randomly-chosen survivors of chemically mutagenised (NTG; 1-methyl-3-nitro-1-nitrosoguanidine) *Schizochytrium* ATCC 20888 screened for variations in fatty acid content. This particular strain was valued for its improved DHA productivity.

A genomic clone described herein as JK1126, isolated from *Schizochytrium* sp. ATCC 20888, comprises, to the best of the present inventors' knowledge, the nucleotide sequence spanning from position 1 to 8730 of SEQ ID NO:1, and encodes the corresponding amino acid sequence of SEQ ID NO:2. Genomic clone pJK1126 (denoted pJK1126 OrfA genomic clone, in the form of an *E. coli* plasmid vector containing "OrfA" gene from *Schizochytrium* ATCC 20888) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7648. The nucleotide sequence of pJK1126 OrfA genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

Two genomic clones described herein as pJK306 OrfA genomic clone and pJK320 OrfA genomic clone, isolated from *Schizochytrium* sp. N230D, together (overlapping clones) comprise, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:1, and encode the amino acid sequence of SEQ ID NO:2. Genomic clone pJK306 (denoted pJK306 OrfA genomic clone, in the form of an *E. coli* plasmid containing 5' portion of OrfA gene from *Schizochytrium* sp. N230D (2.2kB overlap with pJK320)) was deposited with the American Type Culture Collection (ATCC), 10801

University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7641. The nucleotide sequence of pJK306 OrfA genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention. Genomic clone pJK320 (denoted pJK320 OrfA genomic clone, in the form of an *E. coli* plasmid containing 3' portion of OrfA gene from *Schizochytrium* sp. N230D (2.2kB overlap with pJK306)) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7644. The nucleotide sequence of pJK320 OrfA genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in OrfA is a KS domain, also referred to herein as ORFA-KS, and the nucleotide sequence containing the sequence encoding the ORFA-KS domain is represented herein as SEQ ID NO:7 (positions 1-1500 of SEQ ID NO:1). The amino acid sequence containing the ORFA-KS domain is represented herein as SEQ ID NO:8 (positions 1-500 of SEQ ID NO:2). It is noted that the ORFA-KS domain contains an active site motif: DXAC* (*acyl binding site C_{215}). Also, a characteristic motif at the end of the *Schizochytrium* KS region, GFGG, is present in this domain in SEQ ID NO:2 and accordingly, in SEQ ID NO:8.

The second domain in OrfA is a MAT domain, also referred to herein as ORFA-MAT, and the nucleotide sequence containing the sequence encoding the ORFA-MAT domain is represented herein as SEQ ID NO:9 (positions 1723-3000 of SEQ ID NO:1). The amino acid sequence containing the ORFA-MAT domain is represented herein as SEQ ID NO:10 (positions 575-1000 of SEQ ID NO:2). The MAT domain comprises an aspartate at position 93 and a histidine at position 94 (corresponding to positions 667 and 668, respectively, of SEQ ID NO:2). It is noted that the ORFA-MAT domain contains an active site motif: GHS*XG (*acyl binding site S_{706}), represented herein as SEQ ID NO:11.

Domains 3-11 of OrfA are nine tandem ACP domains, also referred to herein as ORFA-ACP (the first domain in the sequence is ORFA-ACP1, the second domain is ORFA-ACP2, the third domain is ORFA-ACP3, etc.). The first ACP domain, ORFA-ACP1, is contained within the nucleotide sequence spanning from about position 3343 to about position 3600 of SEQ ID NO:1 (OrfA). The nucleotide sequence containing the sequence encoding the ORFA-ACP1 domain is represented herein as SEQ ID NO:12

(positions 3343-3600 of SEQ ID NO:1). The amino acid sequence containing the first ACP domain spans from about position 1115 to about position 1200 of SEQ ID NO:2. The amino acid sequence containing the ORFA-ACP1 domain is represented herein as SEQ ID NO:13 (positions 1115-1200 of SEQ ID NO:2). It is noted that the ORFA-ACP1 domain contains an active site motif: LGIDS* (*pantetheine binding motif S_{1157}), represented herein by SEQ ID NO:14.

The nucleotide and amino acid sequences of all nine ACP domains are highly conserved and therefore, the sequence for each domain is not represented herein by an individual sequence identifier. However, based on the information disclosed herein, one of skill in the art can readily determine the sequence containing each of the other eight ACP domains. All nine ACP domains together span a region of OrfA of from about position 3283 to about position 6288 of SEQ ID NO:1, which corresponds to amino acid positions of from about 1095 to about 2096 of SEQ ID NO:2. The nucleotide sequence for the entire ACP region containing all nine domains is represented herein as SEQ ID NO:16. The region represented by SEQ ID NO:16 includes the linker segments between individual ACP domains. The repeat interval for the nine domains is approximately every 330 nucleotides of SEQ ID NO:16 (the actual number of amino acids measured between adjacent active site serines ranges from 104 to 116 amino acids). Each of the nine ACP domains contains a pantetheine binding motif LGIDS* (represented herein by SEQ ID NO:14), wherein S* is the pantetheine binding site serine (S). The pantetheine binding site serine (S) is located near the center of each ACP domain sequence. At each end of the ACP domain region and between each ACP domain is a region that is highly enriched for proline (P) and alanine (A), which is believed to be a linker region. For example, between ACP domains 1 and 2 is the sequence: APAPVKAAAPAAPVASAPAPA, represented herein as SEQ ID NO:15. The locations of the active site serine residues (i.e., the pantetheine binding site) for each of the nine ACP domains, with respect to the amino acid sequence of SEQ ID NO:2, are as follows: ACP1 = S_{1157}; ACP2 = S_{1266}; ACP3 = S_{1377}; ACP4 = S_{1488}; ACP5 = S_{1604}; ACP6 = S_{1715}; ACP7 = S_{1819}; ACP8 = S_{1930}; and ACP9 = S_{2034}. Given that the average size of an ACP domain is about 85 amino acids, excluding the linker, and about 110 amino acids including the linker, with the active site serine being approximately in the center of the domain, one of skill in the art can readily determine the positions of each of the nine ACP domains in OrfA.

Domain 12 in OrfA is a KR domain, also referred to herein as ORFA-KR, and the nucleotide sequence containing the sequence encoding the ORFA-KR domain is represented herein as SEQ ID NO:17 (positions 6598-8730 of SEQ ID NO:1). The amino acid sequence containing the ORFA-KR domain is represented herein as SEQ ID NO:18 (positions 2200-2910 of SEQ ID NO:2). Within the KR domain is a core region with homology to short chain aldehyde-dehydrogenases (KR is a member of this family). This core region spans from about position 7198 to about position 7500 of SEQ ID NO:1, which corresponds to amino acid positions 2400-2500 of SEQ ID NO:2.

Schizochytrium Open Reading Frame B (OrfB):

The complete nucleotide sequence for OrfB is represented herein as SEQ ID NO:3. OrfB is a 6177 nucleotide sequence (not including the stop codon) which encodes a 2059 amino acid sequence, represented herein as SEQ ID NO:4. Within OrfB are four domains: (a) one β-keto acyl-ACP synthase (KS) domain; (b) one chain length factor (CLF) domain; (c) one acyl transferase (AT) domain; and, (d) one enoyl ACP-reductase (ER) domain.

Genomic DNA clones (plasmids) encoding OrfB from both *Schizochytrium* sp. ATCC 20888 and a daughter strain of ATCC 20888, denoted *Schizochytrium* sp., strain N230D, have been isolated and sequenced.

A genomic clone described herein as pJK1129, isolated from *Schizochytrium* sp. ATCC 20888, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:3, and encodes the amino acid sequence of SEQ ID NO:4. Genomic clone pJK1129 (denoted pJK1129 OrfB genomic clone, in the form of an *E. coli* plasmid vector containing "OrfB" gene from *Schizochytrium* ATCC 20888) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7649. The nucleotide sequence of pJK1126 OrfB genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

A genomic clone described herein as pJK324 OrfB genomic clone, isolated from *Schizochytrium* sp. N230D, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:3, and encodes the amino acid sequence of SEQ ID NO:4. Genomic clone pJK324 (denoted pJK324 OrfB genomic clone, in the form of an *E. coli* plasmid containing the OrfB gene sequence from *Schizochytrium* sp. N230D) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC

Accession No. PTA-7643. The nucleotide sequence of pJK324 OrfB genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in OrfB is a KS domain, also referred to herein as ORFB-KS, and the nucleotide sequence containing the sequence encoding the ORFB-KS domain is represented herein as SEQ ID NO:19 (positions 1-1350 of SEQ ID NO:3). The amino acid sequence containing the ORFB-KS domain is represented herein as SEQ ID NO:20 (positions 1-450 of SEQ ID NO:4). This KS domain comprises a valine at position 371 of SEQ ID NO:20 (also position 371 of SEQ ID NO:20). It is noted that the ORFB-KS domain contains an active site motif: DXAC* (*acyl binding site C_{196}). Also, a characteristic motif at the end of this KS region, GFGG, is present in this domain in SEQ ID NO:4 and accordingly, in SEQ ID NO:20.

The second domain in OrfB is a CLF domain, also referred to herein as ORFB-CLF, and the nucleotide sequence containing the sequence encoding the ORFB-CLF domain is represented herein as SEQ ID NO:21 (positions 1378-2700 of SEQ ID NO:3). The amino acid sequence containing the ORFB-CLF domain is represented herein as SEQ ID NO:22 (positions 460-900 of SEQ ID NO:4). It is noted that the ORFB-CLF domain contains a KS active site motif without the acyl-binding cysteine.

The third domain in OrfB is an AT domain, also referred to herein as ORFB-AT, and the nucleotide sequence containing the sequence encoding the ORFB-AT domain is represented herein as SEQ ID NO:23 (positions 2701-4200 of SEQ ID NO:3). The amino acid sequence containing the ORFB-AT domain is represented herein as SEQ ID NO:24 (positions 901-1400 of SEQ ID NO:4). It is noted that the ORFB-AT domain contains an active site motif of GxS*xG (*acyl binding site S_{1140}) that is characteristic of acyltransferse (AT) proteins.

The fourth domain in OrfB is an ER domain, also referred to herein as ORFB-ER, and the nucleotide sequence containing the sequence encoding the ORFB-ER domain is represented herein as SEQ ID NO:25 (positions 4648-6177 of SEQ ID NO:3). The amino acid sequence containing the ORFB-ER domain is represented herein as SEQ ID NO:26 (positions 1550-2059 of SEQ ID NO:4).

Schizochytrium Open Reading Frame C (OrfC):

The complete nucleotide sequence for OrfC is represented herein as SEQ ID NO:5. OrfC is a 4506 nucleotide sequence (not including the stop codon) which encodes a 1502

amino acid sequence, represented herein as SEQ ID NO:6. Within OrfC are three domains: (a) two FabA-like .-hydroxy acyl-ACP dehydrase (DH) domains; and (b) one enoyl ACP-reductase (ER) domain.

Genomic DNA clones (plasmids) encoding OrfC from both *Schizochytrium* sp. ATCC 20888 and a daughter strain of ATCC 20888, denoted *Schizochytrium* sp., strain N230D, have been isolated and sequenced.

A genomic clone described herein as pJK1131, isolated from *Schizochytrium* sp. ATCC 20888, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:5, and encodes the amino acid sequence of SEQ ID NO:6. Genomic clone pJK1131 (denoted pJK1131 OrfC genomic clone, in the form of an *E. coli* plasmid vector containing "OrfC" gene from *Schizochytrium* ATCC 20888) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7650. The nucleotide sequence of pJK1131 OrfC genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

A genomic clone described herein as pBR002 OrfC genomic clone, isolated from *Schizochytrium* sp. N230D, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:5, and encodes the amino acid sequence of SEQ ID NO:6. Genomic clone pBR002 (denoted pBR002 OrfC genomic clone, in the form of an *E. coli* plasmid vector containing the OrfC gene sequence from *Schizochytrium* sp. N230D) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on June 8, 2006, and assigned ATCC Accession No. PTA-7642. The nucleotide sequence of pBR002 OrfC genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in OrfC is a DH domain, also referred to herein as ORFC-DH1. This is one of two DH domains in OrfC, and therefore is designated DH1. The nucleotide sequence containing the sequence encoding the ORFC-DH1 domain is represented herein as SEQ ID NO:27 (positions 1-1350 of SEQ ID NO:5). The amino acid sequence containing the ORFC-DH1 domain is represented herein as SEQ ID NO:28 (positions 1-450 of SEQ ID NO:6).

The second domain in OrfC is a DH domain, also referred to herein as ORFC-DH2. This is the second of two DH domains in OrfC, and therefore is designated DH2.

The nucleotide sequence containing the sequence encoding the ORFC-DH2 domain is represented herein as SEQ ID NO:29 (positions 1351-2847 of SEQ ID NO:5). The amino acid sequence containing the ORFC-DH2 domain is represented herein as SEQ ID NO:30 (positions 451-949 of SEQ ID NO:6). This DH domain comprises the amino acids H-G-I-A-N-P-T-F-V-H-A-P-G-K-I (positions 876-890 of SEQ ID NO:6) at positions 426-440 of SEQ ID NO:30.

The third domain in OrfC is an ER domain, also referred to herein as ORFC-ER, and the nucleotide sequence containing the sequence encoding the ORFC-ER domain is represented herein as SEQ ID NO:31 (positions 2995-4506 of SEQ ID NO:5). The amino acid sequence containing the ORFC-ER domain is represented herein as SEQ ID NO:32 (positions 999-1502 of SEQ ID NO:6).

***Thraustochytrium* PUFA PKS System**

There are three open reading frames that form the core *Thraustochytrium* 23B PUFA PKS system. The domain organization is the same as that of *Schizochytrium* with the exception that the *Th.* 23B Orf A has 8 adjacent ACP domains, while *Schizochytrium* Orf A has 9 adjacent ACP domains. The domain structure of each open reading frame is as follows.

Thraustochytrium 23B Open Reading Frame A (OrfA):

The complete nucleotide sequence for *Th.* 23B OrfA is represented herein as SEQ ID NO:38. *Th.* 23B OrfA is a 8433 nucleotide sequence (not including the stop codon) which encodes a 2811 amino acid sequence, represented herein as SEQ ID NO:39. SEQ ID NO:38 encodes the following domains in *Th.* 23B OrfA: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one malonyl-CoA:ACP acyltransferase (MAT) domain; (c) eight acyl carrier protein (ACP) domains; and (d) one β-ketoacyl-ACP reductase (KR) domain.

Two genomic clones described herein as Th23BOrfA_pBR812.1 and Th23BOrfA_pBR811 (OrfA genomic clones), isolated from *Thraustochytrium* 23B, together (overlapping clones) comprise, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:38, and encodes the amino acid sequence of SEQ ID NO:39. Genomic clone Th23BOrfA_pBR812.1 (denoted Th23BOrfA_pBR812.1 genomic clone, in the form of an *E. coli* plasmid vector containing the OrfA gene sequence from *Thraustochytrium 23B*) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on

March 1, 2007, and assigned ATCC Accession No. PTA-8232. The nucleotide sequence of Th23BOrfA_pBR812.1, an OrfA genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention. Genomic clone Th23BOrfA_pBR811 (denoted Th23BOrfA_pBR811 genomic clone, in the form of an *E. coli* plasmid vector containing the OrfA gene sequence from *Thraustochytrium 23B*) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8231. The nucleotide sequence of Th23BOrfA_pBR811, an OrfA genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in *Th.* 23B OrfA is a KS domain, also referred to herein as *Th.* 23B OrfA-KS, and is contained within the nucleotide sequence spanning from about position 1 to about position 1500 of SEQ ID NO:38, represented herein as SEQ ID NO:40. The amino acid sequence containing the *Th.* 23B KS domain is a region of SEQ ID NO:39 spanning from about position 1 to about position 500 of SEQ ID NO:39, represented herein as SEQ ID NO:41. This region of SEQ ID NO:39 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from position 1 to about position 450 of SEQ ID NO:39 (also positions 1 to about 450 of SEQ ID NO:41). It is noted that the *Th.* 23B OrfA-KS domain contains an active site motif: DXAC* (*acyl binding site C_{207}). Also, a characteristic motif at the end of the *Th.* 23B KS region, GFGG, is present in positions 453-456 of SEQ ID NO:39 (also positions 453-456 of SEQ ID NO:41).

The second domain in *Th.* 23B OrfA is a MAT domain, also referred to herein as *Th.* 23B OrfA-MAT, and is contained within the nucleotide sequence spanning from between about position 1503 and about position 3000 of SEQ ID NO:38, represented herein as SEQ ID NO:42. The amino acid sequence containing the *Th.* 23B MAT domain is a region of SEQ ID NO:39 spanning from about position 501 to about position 1000, represented herein by SEQ ID NO:43. This region of SEQ ID NO:39 has a Pfam match to FabD (malonyl-CoA:ACP acyltransferase) spanning from about position 580 to about position 900 of SEQ ID NO:39 (positions 80-400 of SEQ ID NO:43). It is noted that the *Th.* 23B OrfA-MAT domain contains an active site motif: GHS*XG (*acyl binding site S_{697}), represented by positions 695-699 of SEQ ID NO:39.

Domains 3-10 of *Th.* 23B OrfA are eight tandem ACP domains, also referred to herein as *Th.* 23B OrfA-ACP (the first domain in the sequence is OrfA-ACP1, the second

domain is OrfA-ACP2, the third domain is OrfA-ACP3, etc.). The first *Th.* 23B ACP domain, *Th.* 23B OrfA-ACP1, is contained within the nucleotide sequence spanning from about position 3205 to about position 3555 of SEQ ID NO:38 (OrfA), represented herein as SEQ ID NO:44. The amino acid sequence containing the first *Th.* 23B ACP domain is a region of SEQ ID NO:39 spanning from about position 1069 to about position 1185 of SEQ ID NO:39, represented herein by SEQ ID NO:45.

The eight ACP domains in *Th.* 23B OrfA are adjacent to one another and can be identified by the presence of the phosphopantetheine binding site motif, LGXDS* (represented by SEQ ID NO:46), wherein the S* is the phosphopantetheine attachment site. The amino acid position of each of the eight S* sites, with reference to SEQ ID NO:39, are 1128 (ACP1), 1244 (ACP2), 1360 (ACP3), 1476 (ACP4), 1592 (ACP5), 1708 (ACP6), 1824 (ACP7) and 1940 (ACP8). The nucleotide and amino acid sequences of all eight *Th.* 23B ACP domains are highly conserved and therefore, the sequence for each domain is not represented herein by an individual sequence identifier. However, based on the information disclosed herein, one of skill in the art can readily determine the sequence containing each of the other seven ACP domains in SEQ ID NO:38 and SEQ ID NO:39.

All eight *Th.* 23B ACP domains together span a region of *Th.* 23B OrfA of from about position 3205 to about position 5994 of SEQ ID NO:38, which corresponds to amino acid positions of from about 1069 to about 1998 of SEQ ID NO:39. The nucleotide sequence for the entire ACP region containing all eight domains is represented herein as SEQ ID NO:47. SEQ ID NO:47 encodes an amino acid sequence represented herein by SEQ ID NO:48. SEQ ID NO:48 includes the linker segments between individual ACP domains. The repeat interval for the eight domains is approximately every 116 amino acids of SEQ ID NO:48, and each domain can be considered to consist of about 116 amino acids centered on the active site motif (described above).

The last domain in *Th.* 23B OrfA is a KR domain, also referred to herein as *Th.* 23B OrfA-KR, which is contained within the nucleotide sequence spanning from between about position 6001 to about position 8433 of SEQ ID NO:38, represented herein by SEQ ID NO:49. The amino acid sequence containing the *Th.* 23B KR domain is a region of SEQ ID NO:39 spanning from about position 2001 to about position 2811 of SEQ ID NO:39, represented herein by SEQ ID NO:50. This region of SEQ ID NO:39 has a Pfam match to FabG (β-ketoacyl-ACP reductase) spanning from about position 2300 to about 2550 of SEQ ID NO:39 (positions 300-550 of SEQ ID NO:50).

Thraustochytrium. 23B Open Reading Frame B (OrfB):

The complete nucleotide sequence for *Th.* 23B OrfB is represented herein as SEQ ID NO:51, which is a 5805 nucleotide sequence (not including the stop codon) that encodes a 1935 amino acid sequence, represented herein as SEQ ID NO:52. SEQ ID NO:51 encodes the following domains in *Th.* 23B OrfB: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one chain length factor (CLF) domain; (c) one acyltransferase (AT) domain; and, (d) one enoyl-ACP reductase (ER) domain.

A genomic clone described herein as Th23BOrfB_pBR800 (OrfB genomic clone), isolated from *Thraustochytrium* 23B, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:51, and encodes the amino acid sequence of SEQ ID NO:52. Genomic clone Th23BOrfB_pBR800 (denoted Th23BOrfB_pBR800 genomic clone, in the form of an *E. coli* plasmid vector containing the OrfB gene sequence from *Thraustochytrium 23B*) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8227. The nucleotide sequence of Th23BOrfB_pBR800, an OrfB genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in the *Th.* 23B OrfB is a KS domain, also referred to herein as *Th.* 23B OrfB-KS, which is contained within the nucleotide sequence spanning from between about position 1 and about position 1500 of SEQ ID NO:51 (*Th.* 23B OrfB), represented herein as SEQ ID NO:53. The amino acid sequence containing the *Th.* 23B KS domain is a region of SEQ ID NO: 52 spanning from about position 1 to about position 500 of SEQ ID NO:52, represented herein as SEQ ID NO:54. This region of SEQ ID NO:52 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from about position 1 to about position 450 (positions 1-450 of SEQ ID NO:54). It is noted that the *Th.* 23B OrfB-KS domain contains an active site motif: DXAC*, where C* is the site of acyl group attachment and wherein the C* is at position 201 of SEQ ID NO:52. Also, a characteristic motif at the end of the KS region, GFGG is present in amino acid positions 434-437 of SEQ ID NO:52.

The second domain in *Th.* 23B OrfB is a CLF domain, also referred to herein as *Th.* 23B OrfB-CLF, which is contained within the nucleotide sequence spanning from between about position 1501 and about position 3000 of SEQ ID NO:51 (OrfB), represented herein as SEQ ID NO:55. The amino acid sequence containing the CLF

domain is a region of SEQ ID NO: 52 spanning from about position 501 to about position 1000 of SEQ ID NO:52, represented herein as SEQ ID NO:56. This region of SEQ ID NO:52 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from about position 550 to about position 910 (positions 50-410 of SEQ ID NO:56). Although CLF has homology to KS proteins, it lacks an active site cysteine to which the acyl group is attached in KS proteins.

The third domain in *Th.* 23B OrfB is an AT domain, also referred to herein as *Th.* 23B OrfB-AT, which is contained within the nucleotide sequence spanning from between about position 3001 and about position 4500 of SEQ ID NO:51 (*Th.* 23B OrfB), represented herein as SEQ ID NO:58. The amino acid sequence containing the *Th.* 23B AT domain is a region of SEQ ID NO: 52 spanning from about position 1001 to about position 1500 of SEQ ID NO:52, represented herein as SEQ ID NO:58. This region of SEQ ID NO:52 has a Pfam match to FabD (malonyl-CoA:ACP acyltransferase) spanning from about position 1100 to about position 1375 (positions 100-375 of SEQ ID NO:58). Although this AT domain of the PUFA synthases has homology to MAT proteins, it lacks the extended motif of the MAT (key arginine and glutamine residues) and it is not thought to be involved in malonyl-CoA transfers. The GXS*XG motif of acyltransferases is present, with the S* being the site of acyl attachment and located at position 1123 with respect to SEQ ID NO:52.

The fourth domain in *Th.* 23B OrfB is an ER domain, also referred to herein as *Th.* 23B OrfB-ER, which is contained within the nucleotide sequence spanning from between about position 4501 and about position 5805 of SEQ ID NO:51 (OrfB), represented herein as SEQ ID NO:59. The amino acid sequence containing the *Th.* 23B ER domain is a region of SEQ ID NO: 52 spanning from about position 1501 to about position 1935 of SEQ ID NO:52, represented herein as SEQ ID NO:60. This region of SEQ ID NO:52 has a Pfam match to a family of dioxygenases related to 2-nitropropane dioxygenases spanning from about position 1501 to about position 1810 (positions 1-310 of SEQ ID NO:60). That this domain functions as an ER can be further predicted due to homology to a newly characterized ER enzyme from *Streptococcus pneumoniae*.

Thraustochytrium. 23B Open Reading Frame C (OrfC):

The complete nucleotide sequence for *Th.* 23B OrfC is represented herein as SEQ ID NO:61, which is a 4410 nucleotide sequence (not including the stop codon) that encodes a 1470 amino acid sequence, represented herein as SEQ ID NO:62. SEQ ID

NO:61 encodes the following domains in *Th.* 23B OrfC: (a) two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains, both with homology to the FabA protein (an enzyme that catalyzes the synthesis of *trans*-2-decenoyl-ACP and the reversible isomerization of this product to *cis*-3-decenoyl-ACP); and (b) one enoyl-ACP reductase (ER) domain with high homology to the ER domain of *Schizochytrium* OrfB.

A genomic clone described herein as Th23BOrfC_pBR709A (OrfC genomic clone), isolated from *Thraustochytrium* 23B, comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:61, and encodes the amino acid sequence of SEQ ID NO:62. Genomic clone Th23BOrfC_pBR709A (denoted Th23BOrfC_pBR709A genomic clone, in the form of an *E. coli* plasmid vector containing the OrfC gene sequence from *Thraustochytrium 23B*) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8228. The nucleotide sequence of Th23BOrfC_pBR709A, an OrfC genomic clone, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

The first domain in *Th.* 23B OrfC is a DH domain, also referred to herein as *Th.* 23B OrfC-DH1, which is contained within the nucleotide sequence spanning from between about position 1 to about position 1500 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:63. The amino acid sequence containing the *Th.* 23B DH1 domain is a region of SEQ ID NO: 62 spanning from about position 1 to about position 500 of SEQ ID NO:62, represented herein as SEQ ID NO:64. This region of SEQ ID NO:62 has a Pfam match to FabA, as mentioned above, spanning from about position 275 to about position 400 (positions 275-400 of SEQ ID NO:64).

The second domain in *Th.* 23B OrfC is also a DH domain, also referred to herein as *Th.* 23B OrfC-DH2, which is contained within the nucleotide sequence spanning from between about position 1501 to about 3000 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:65. The amino acid sequence containing the *Th.* 23B DH2 domain is a region of SEQ ID NO: 62 spanning from about position 501 to about position 1000 of SEQ ID NO:62, represented herein as SEQ ID NO:66. This region of SEQ ID NO:62 has a Pfam match to FabA, as mentioned above, spanning from about position 800 to about position 925 (positions 300-425 of SEQ ID NO:66).

The third domain in *Th.* 23B OrfC is an ER domain, also referred to herein as *Th.* 23B OrfC-ER, which is contained within the nucleotide sequence spanning from between

about position 3001 to about position 4410 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:67. The amino acid sequence containing the *Th.* 23B ER domain is a region of SEQ ID NO: 62 spanning from about position 1001 to about position 1470 of SEQ ID NO:62, represented herein as SEQ ID NO:68. This region of SEQ ID NO:62 has a Pfam match to the dioxygenases related to 2-nitropropane dioxygenases, as mentioned above, spanning from about position 1025 to about position 1320 (positions 25-320 of SEQ ID NO:68). This domain function as an ER can also be predicted due to homology to a newly characterized ER enzyme from *Streptococcus pneumoniae*.

Synthetic, Codon-Optimized Constructs

The invention also encompasses resynthesized versions of any of the nucleic acid sequences described herein, primarily having optimized codon usage for a heterologous organism (heterologous host), wherein the encoded amino acid sequence is not changed with reference to the natural, wild-type, or source amino acid sequence. The present inventors have discovered that resynthesizing nucleic acid sequences for optimal codon usage is an effective way to improve PUFA production in a heterologous host that is transformed with nucleic acid molecules from a PUFA PKS system. Resynthesis of all nucleic acid molecules in a PUFA PKS system is not necessarily required for optimal expression and PUFA production in a heterologous host. Indeed, the inventors have found that resynthesis of only some of the nucleic acid molecules is sufficient to improve PUFA production. For example, while resynthesis of *Schizochytrium* Orfs A and B improved PUFA synthase expression and PUFA production in yeast, use of the native *Schizochytrium* OrfC and native *Nostoc* HetI PPTase were sufficient. Moreover, codon optimization of a construct for use in one heterologous host may also be useful for improving the PUFA production in a different heterologous host (*e.g.,* optimization of codon usage of an OrfC-encoding sequence from *Thraustochytrium* for use in *Schizochytrium* may also be effective for boosting PUFA production in another heterologous host organism, such as plants).

In addition, the use of synthetic, codon-optimized constructs can be useful in the production of chimeric PUFA PKS constructs and/or chimeric PUFA PKS systems, where a domain or protein from one PUFA PKS system (*e.g.,* from a first organism) is introduced into a second PUFA PKS system (*e.g.,* from a second organism). In such systems, not only can the PUFA profile be manipulated (*e.g.,* by the use of the chimeric constructs and/or chimeric PUFA PKS systems), but the PUFA production can also be

improved by the use of synthetic, codon-optimized chimeric constructs. Indeed, the combination of the two concepts (chimeras and codon optimization) may produce a synergistic result with respect to PUFA profiles and/or PUFA production. Chimeric systems containing some sequences that are codon-optimized for the host and some that are not codon-optimized for the host are included in the invention.

Certain codon-optimized sequences are described below by way of example. Other codon-optimized sequences will be apparent to those of skill in the art following this description.

sOrfA

SEQ ID NO:35, denoted sOrfA, represents the nucleic acid sequence encoding OrfA from *Schizochytrium* (SEQ ID NO:1) that has been resynthesized for optimized codon usage in yeast. SEQ ID NO:1 and SEQ ID NO:35 each encode SEQ ID NO:2.

sOrfB

SEQ ID NO:36, denoted sOrfB, represents the nucleic acid sequence encoding OrfB from *Schizochytrium* (SEQ ID NO:3) that has been resynthesized for optimized codon usage in yeast. SEQ ID NO:3 and SEQ ID NO:36 each encode SEQ ID NO:4.

*OrfB**

SEQ ID NO:37, denoted OrfB* (pJK962), represents a nucleic acid sequence encoding OrfB from *Schizochytrium* (SEQ ID NO:4) that has been resynthesized within a portion of SEQ ID NO:3 (nucleotide sequence encoding SEQ ID NO:4) for use in plant cells, and that was derived from a very similar sequence initially developed for optimized codon usage in *E. coli*, also referred to as OrfB* (pJK780), which is described below. OrfB* in both forms (for *E. coli* and for plants) is identical to SEQ ID NO:3 with the exception of a resynthesized BspHI (nucleotide 4415 of SEQ ID NO:3) to a SacII fragment (unique site in SEQ ID NO:3). Both versions (*E. coli* and plant) have two other codon modifications near the start of the gene as compared with the original genomic sequence of orfB (SEQ ID NO:3). First, the fourth codon, arginine (R), was changed from CGG in the genomic sequence to CGC in orfB*. Second, the fifth codon, asparagine (N), was changed from AAT in the genomic sequence to AAC in orf B*. In order to facilitate cloning of this gene into the plant vectors to create SEQ ID NO:37, a PstI site (CTGCAG) was also engineered into the *E. coli* orfB* sequence 20 bases from the start of the gene. This change did not alter the amino acid sequence of the encoded protein. Both SEQ ID

NO:37 and SEQ ID NO:3 (as well as the OrfB* form for *E. coli*, described in SEQ ID NO:69 below) encode SEQ ID NO:4.

SEQ ID NO:69, denoted OrfB* (pJK780), represents a nucleic acid sequence encoding OrfB from *Schizochytrium* (SEQ ID NO:4) that has been resynthesized within a portion of SEQ ID NO:3 (nucleotide sequence encoding SEQ ID NO:4) for use in *E. coli*. The sequence of the OrfB* construct in both forms (for *E. coli* and for plants) has been described above. SEQ ID NO:69 and SEQ ID NO:3 encode SEQ ID NO:4.

The plasmid described herein as OrfB*_pJK780 comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:69, and encodes the amino acid sequence of SEQ ID NO:4. Plasmid OrfB*_pJK780 (denoted OrfB*_pJK780 clone, in the form of an *E. coli* plasmid vector) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8225. The nucleotide sequence of OrfB*_pJK780 and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

pThOrfC-synPS

SEQ ID NO:70 represents a nucleic acid sequence encoding a *Thraustochytrium* 23B OrfC (SEQ ID NO:61, encoding SEQ ID NO:62) that has been resynthesized for optimized codon usage in *Schizochytrium*. Positions 2000-6412 of SEQ ID NO:70 represents the coding region for the *Thraustochytrium* 23B OrfC protein (including stop codon). Positions 1-1999 and 6413-8394 of SEQ ID NO:70 represent upstream and downstream *Schizochytrium* OrfC sequences (non-coding regions), respectively. The construction of the plasmid containing SEQ ID NO:70, denoted pThOrfC-synPS, is described in detail in Example 1. SEQ ID NO:70 and SEQ ID NO:61 each encode SEQ ID NO:62. pThOrfC-syn PS is designed to exactly replace the coding region (CDS) of *Schizochytrium* orfC (SEQ ID NO;5) with the coding region for the *Thraustochytrium* 23B orfC, resynthesized as discussed above (SEQ ID NO:70). The production and use of organisms that have been transformed with this construct are described in detail below and in the Examples.

The plasmid described above as pThOrfC-synPS comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:70, and encodes the corresponding amino acid sequence of SEQ ID NO:62. Plasmid pThOrfC-synPS (denoted pThOrfC-synPS, in the form of an *E. coli* plasmid vector containing a "perfect stitch"

synthetic Thraustochytrium 23B PUFA PKS OrfC codon optimized for expression in Schizochytrium or other heterologous hosts) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8229. The nucleotide sequence of pThOrfC-synPS, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

pDD26

SEQ ID NO:71 represents a nucleic acid sequence encoding a *Thraustochytrium* 23B OrfA (SEQ ID NO:38, encoding SEQ ID NO:39) that has been resynthesized for optimized codon usage in *Schizochytrium*. Positions 2044-10479 of SEQ ID NO:71 represents the coding region for the *Thraustochytrium* 23B OrfA protein (including stop codon). Positions 1-2043 and 10480-12495 of SEQ ID NO:71 represent upstream and downstream *Schizochytrium* OrfA sequences (non-coding regions), respectively. The construction of the plasmid containing SEQ ID NO:71, denoted pDD26, is described in detail in Example 8. SEQ ID NO:71 and SEQ ID NO:38 each encode SEQ ID NO:39. pDD26 is designed to exactly replace the coding region (CDS) of *Schizochytrium* orfA (SEQ ID NO:1) with the coding region for the *Thraustochytrium* 23B orfC, resynthesized as discussed above (SEQ ID NO:71). The production and use of organisms that have been transformed with this construct are described in detail below and in the Examples.

The plasmid described above as pDD26 comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:71, and encodes the corresponding amino acid sequence of SEQ ID NO:39. Plasmid pDD26 (denoted pDD26, in the form of an *E. coli* plasmid vector) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on May 8, 2007, and assigned ATCC Accession No. PTA-8411. The nucleotide sequence of pDD26, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

pDD32

SEQ ID NO:72 represents a nucleic acid sequence encoding a *Thraustochytrium* 23B OrfB (SEQ ID NO:51, encoding SEQ ID NO:52) that has been resynthesized for optimized codon usage in *Schizochytrium*. Positions 1452-7259 of SEQ ID NO:72 represent the coding region for the *Thraustochytrium* 23B OrfB protein (including stop codon). Positions 1-1451 and 7260-8647 of SEQ ID NO:72 represent upstream and

downstream *Schizochytrium* OrfB sequences (non-coding regions), respectively. The construction of the plasmid containing SEQ ID NO:72, denoted pDD32, is described in detail in Example 8. SEQ ID NO:72 and SEQ ID NO:51 each encode SEQ ID NO:52. pDD32 is designed to exactly replace the coding region (CDS) of *Schizochytrium* orfB (SEQ ID NO:3) with the coding region for the *Thraustochytrium* 23B orfC, resynthesized as discussed above (SEQ ID NO:72). The production and use of organisms that have been transformed with this construct are described in detail below and in the Examples.

The plasmid described above as pDD32 comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:72, and encodes the corresponding amino acid sequence of SEQ ID NO:52. Plasmid pDD32 (denoted pDD32, in the form of an *E. coli* plasmid vector) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on May 8, 2007, and assigned ATCC Accession No. PTA-8412. The nucleotide sequence of pDD32, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

Chimeric PUFA PKS Constructs

The invention also encompasses chimeric constructs using portions of two or more different PUFA PKS nucleic acid sequences, such as those described herein, to produce chimeric PUFA PKS proteins. The present inventors demonstrate herein in several different examples that by "mixing and matching" domains or portions of PUFA PKS proteins from different organisms (*i.e.,* creating chimeric PUFA PKS proteins comprised of domains or polypeptides from two or more different organisms), the profile of the PUFAs produced by an organism expressing a PUFA PKS system containing such chimeric proteins can be modified, as compared to a native (naturally occurring) PUFA PKS system. For example, the present inventors describe herein the use of the DH2 domain from a *Thraustochytrium* PUFA PKS system in the OrfC protein of a *Schizochytrium* protein, so that the resulting chimeric OrfC protein contains the DH1 and ER domains from *Schizochytrium*, and the DH2 domain from *Thraustochytrium*. The chimeric construct is further modified by the use of a codon-optimized (for *Schizochytrium*) *Thraustochytrium* DH2 domain in one construct, and a native *Thraustochytrium* DH2 domain in another construct, which demonstrates the flexibility and effects of the various modifications described herein.

Certain chimeric constructs are described below by way of example. Other chimeric constructs will be apparent to those of skill in the art following this description.

pDS49

SEQ ID NO:73 represents a nucleic acid sequence encoding a chimeric protein comprising a *Schizochytrium* OrfC protein (SEQ ID NO:6) in which the DH2 domain (SEQ ID NO:30) has been replaced with the DH2 domain (sequence including SEQ ID NO:66) from *Thraustochytrium* 23B OrfC (SEQ ID NO:62). In this chimeric construct, the DH2-encoding sequence from *Thraustochytrium* is the native (non-codon-optimized) sequence. The construction of the plasmid containing SEQ ID NO:73, denoted pDS49, is described in detail in Example 2. The *Schizochytrium* OrfC upstream and downstream non-coding sequences that flank SEQ ID NO:73 in pDS49 are the same as those described above with respect to SEQ ID NO:70 (not represented in SEQ ID NO:73). SEQ ID NO:73 encodes an amino acid sequence of SEQ ID NO:74. Referring to SEQ ID NO:74, the chimeric OrfC polypeptide is 1493 amino acid residues in length. The DH2 region, defined as amino acids 516–1041 of SEQ ID NO:74, consists of the amino acid sequence of the DH2 region of the *Th.*23B OrfC protein, that is, amino acids 491–1016 of SEQ ID NO:62, which includes all of SEQ ID NO:66 and some flanking amino acid sequence from SEQ ID NO:62. With respect to the remainder of the chimeric OrfC amino acid sequence, residues 1–515 and 1042–1493 of SEQ ID NO:74 are identical to *Schizochytrium* OrfC residues 1–515 and 1051–1502 of SEQ ID NO:6, respectively. The production and use of organisms that have been transformed with this construct are described in detail below and in the Examples.

The plasmid described above as pDS49 comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:73, and encodes the corresponding amino acid sequence of SEQ ID NO:74. Plasmid pDS49 (denoted pDS49, in the form of an *E. coli* plasmid vector) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8230. The nucleotide sequence of pDS49, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

pDD24

SEQ ID NO:75 represents another nucleic acid sequence encoding a chimeric protein comprising a *Schizochytrium* OrfC protein (SEQ ID NO:6) in which the DH2

domain (SEQ ID NO:30) has been replaced with the DH2 domain (sequence including SEQ ID NO:66) from *Thraustochytrium* 23B OrfC (SEQ ID NO:62). In this chimeric construct, the DH2-encoding sequence from *Thraustochytrium* is a codon-optimized sequence for use in *Schizochytrium*. The construction of the plasmid containing SEQ ID NO:75, denoted pDD24, is described in detail in Example 3. The *Schizochytrium* OrfC upstream and downstream non-coding sequences that flank SEQ ID NO:75 in pDD24 are the same as those described above with respect to SEQ ID NO:70 (not represented in SEQ ID NO:75). SEQ ID NO:75 encodes an amino acid sequence of SEQ ID NO:74. SEQ ID NO:74 has been described in detail above with respect to SEQ ID NO:73, which also encodes SEQ ID NO:74. However, in this construct, as discussed above the nucleotide sequence encoding amino acids 516–1041 of SEQ ID NO:74 was derived from the "synthetic gene sequence" for OrfC of *Thraustochytrium*.23B that is contained in plasmid pThOrfC-synPS (see Example 1 and SEQ ID NO:70) and which employs codons that are preferred for gene expression in *Schizochytrium*. The production and use of organisms that have been transformed with this construct are described in detail below and in the Examples.

The plasmid described above as pDD24 comprises, to the best of the present inventors' knowledge, the nucleotide sequence of SEQ ID NO:75, and encodes the corresponding amino acid sequence of SEQ ID NO:74. Plasmid pDD24 (denoted pDD24, in the form of an *E. coli* plasmid vector) was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA on March 1, 2007, and assigned ATCC Accession No. PTA-8226. The nucleotide sequence of pDD24, and the amino acid sequence encoded by this plasmid are encompassed by the present invention.

Chimeric PUFA PKS Systems

In addition to the use of codon-optimization and chimeric constructs described above, the invention includes the production and use of chimeric PUFA PKS systems. Chimeric PUFA PKS systems include the use of the chimeric constructs described above, where a chimeric PUFA PKS protein is created and used in a PUFA PKS system, but such systems also encompass PUFA PKS systems where one or more entire protein or proteins from one or more PUFA PKS system(s) are exchanged for or added to the corresponding entire protein or proteins from another PUFA PKS system, such that the resulting PUFA PKS system comprises proteins from two or more different PUFA PKS systems. Such

systems can also include the use of chimeric proteins, as described above (*e.g.,* chimeric proteins, and substitutions of whole proteins). For example, the construct described above as pTh23B_synPS (comprising a *Thraustochytrium* 23B OrfC-encoding sequence, optimized for *Schizochytrium* codon usage) can be substituted into a *Schizochytrium* PUFA PKS system to perfectly replace the native *Schizochytrium* OrfC-encoding sequence, thereby creating a chimeric PUFA PKS system. As another example, the native *Thraustochytrium* 23B OrfC-encoding sequence (not codon-optimized) can be can be substituted into a *Schizochytrium* PUFA PKS system to perfectly replace the native *Schizochytrium* OrfC-encoding sequence, thereby creating another chimeric PUFA PKS system. As yet another example, the native *Thraustochytrium* 23B OrfA- and OrfC-encoding sequences (codon-optimized, or not) can be substituted into a *Schizochytrium* PUFA PKS system to perfectly replace the native *Schizochytrium* OrfA- and OrfC-encoding sequences, respectively, thereby creating yet another chimeric PUFA PKS system. These and other chimeric PUFA PKS systems are described in the Examples below. Included in the Examples are *Schizochytrium* hosts expressing chimeric PUFA PKS systems comprised of: (1) *Schizochytrium* (*S*) OrfA, *S*OrfB, and *Thraustochytrium* (*Th*) OrfC; (2) *S*OrfA, *Th*OrfB, and *S*OrfC; (3) *Th*OrfA, *S*OrfB, and *S*OrfC; (4) *S*OrfA, *Th*OrfB, and *Th*OrfC; (5) *Th*OrfA, *S*OrfB, and *Th*OrfC; (6) *Th*OrfA, *Th*OrfB, and *S*OrfC; and (7) *Th*OrfA, *Th*OrfB, and *Th*OrfC.

Based on the discussion and exemplary experiments provided herein, it is now possible to improve and/or modify PUFA production by selected resynthesis of PUFA PKS nucleic acid molecules for host codon usage, and/or the use of chimeric PUFA PKS constructs and/or chimeric PUFA PKS systems in various host organisms, including in host organisms that do not endogenously have a PUFA PKS system for the production of PUFAs.

Phosphopantetheinyl transferase (PPTase)

According to the present invention, a PUFA PKS system for production and/or accumulation of PUFAs in a heterologous host or improved production and/or accumulation of PUFAs in an endogenous host may make use of various accessory proteins, which are defined herein as proteins that are not considered to be part of the core PUFA PKS system as described above (*i.e.,* not part of the PUFA synthase enzyme complex itself), but which may be, or are, necessary for PUFA production or at least for

efficient PUFA production using the core PUFA synthase enzyme complex of the present invention.

In order to produce PUFAs, a PUFA PKS system must work with an accessory protein that transfers a 4'-phosphopantetheinyl moiety from coenzyme A to the acyl carrier protein (ACP) domain(s). Therefore, a PUFA PKS system can be considered to include at least one 4'-phosphopantetheinyl transferase (PPTase) domain, or such a domain can be considered to be an accessory domain or protein to the PUFA PKS system. Structural and functional characteristics of PPTases have been described in detail, for example, in U.S. Patent Application Publication No. 20020194641; U.S. Patent Application Publication No. 20040235127; and U.S. Patent Application Publication No. 20050100995.

According to the present invention, a domain or protein having 4'-phosphopantetheinyl transferase (PPTase) biological activity (function) is characterized as the enzyme that transfers a 4'-phosphopantetheinyl moiety from Coenzyme A to the acyl carrier protein (ACP). This transfer to an invariant serine reside of the ACP activates the inactive apo-form to the holo-form. In both polyketide and fatty acid synthesis, the phosphopantetheine group forms thioesters with the growing acyl chains. The PPTases are a family of enzymes that have been well characterized in fatty acid synthesis, polyketide synthesis, and non-ribosomal peptide synthesis. The sequences of many PPTases are known, and crystal structures have been determined (e.g., Reuter K, Mofid MR, Marahiel MA, Ficner R. "Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily" EMBO J. 1999 Dec 1;18(23):6823-31) as well as mutational analysis of amino acid residues important for activity (Mofid MR, Finking R, Essen LO, Marahiel MA. "Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from Bacillus subtilis: carrier protein recognition and reaction mechanism" Biochemistry. 2004 Apr 13;43(14):4128-36). These invariant and highly conserved amino acids in PPTases are contained within the pfaE ORFs from both *Shewanella* strains described above.

One heterologous PPTase which has been demonstrated previously to recognize the OrfA ACP domains described herein as substrates is the Het I protein of *Nostoc* sp. PCC 7120 (formerly called *Anabaena* sp. PCC 7120). Het I is present in a cluster of genes in *Nostoc* known to be responsible for the synthesis of long chain hydroxy-fatty acids that are a component of a glyco-lipid layer present in heterocysts of that organism (Black and

Wolk, 1994, *J. Bacteriol.* 176, 2282-2292; Campbell et al., 1997, *Arch. Microbiol.* 167, 251-258). Het I is likely to activate the ACP domains of a protein, Hgl E, present in that cluster. The two ACP domains of Hgl E have a high degree of sequence homology to the ACP domains found in *Schizochytrium* Orf A. SEQ ID NO:34 represents the amino acid sequence of the Nostoc Het I protein, and is a functional PPTase that can be used with a PUFA PKS system described herein, including the PUFA PKS systems from *Schizochytrium* and *Thraustochytrium*. SEQ ID NO:34 is encoded by SEQ ID NO:33. The endogenous start codon of Het I has not been identified (there is no methionine present in the putative protein). There are several potential alternative start codons (e.g., TTG and ATT) near the 5' end of the open reading frame. No methionine codons (ATG) are present in the sequence. However, the construction of a Het I expression construct was completed using PCR to replace the furthest 5' potential alternative start codon (TTG) with a methionine codon (ATG, as part of an NdeI restriction enzyme recognition site), and introducing an XhoI site at the 3' end of the coding sequence, and the encoded PPTase (SEQ ID NO:34) has been shown to be functional.

Another heterologous PPTase which has been demonstrated previously to recognize the OrfA ACP domains described herein as substrates is sfp, derived from *Bacillus subtilis*. Sfp has been well characterized, and is widely used due to its ability to recognize a broad range of substrates. Based on published sequence information (Nakana, et al., 1992, *Molecular and General Genetics* 232: 313-321), an expression vector was previously produced for sfp by cloning the coding region, along with defined up- and downstream flanking DNA sequences, into a pACYC-184 cloning vector. This construct encodes a functional PPTase as demonstrated by its ability to be co-expressed with *Schizochytrium* Orfs A, B*, and C in *E. coli* which, under appropriate conditions, resulted in the accumulation of DHA in those cells (see U.S. Patent Application Publication No. 20040235127).

When genetically modifying organisms (*e.g.,* microorganisms or plants) to express a PUFA PKS system according to the present invention, some host organisms may endogenously express accessory proteins that are needed to work with the PUFA PKS to produce PUFAs (*e.g.,* PPTases). However, some organisms may be transformed with nucleic acid molecules encoding one or more accessory proteins described herein to enable and/or to enhance production of PUFAs by the organism, even if the organism endogenously produces a homologous accessory protein (*i.e.,* some heterologous

accessory proteins may operate more effectively or efficiently with the transformed PUFA synthase proteins than the host cells' endogenous accessory protein). In one embodiment, such an accessory protein includes an accessory PPTase.

One embodiment of the present invention relates to an isolated nucleic acid molecule comprising a nucleic acid sequence from a PUFA PKS system, a homologue thereof, a fragment thereof, and/or a nucleic acid sequence that is complementary to any of such nucleic acid sequences. In one aspect, the present invention relates to an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of: (a) a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:39, SEQ ID NO:52, SEQ ID NO:62, and biologically active fragments thereof; (b) a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of: SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, and biologically active fragments thereof; (c) a nucleic acid sequence encoding an amino acid sequence that is at least about 60% identical to at least 500 consecutive amino acids of any of the amino acid sequences of (a), wherein the amino acid sequence has a biological activity of at least one, two, three or more domains of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) a nucleic acid sequence encoding an amino acid sequence that is at least about 60% identical to any of the amino acid sequences of (b), wherein said amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; or (e) a nucleic acid sequence that is fully complementary to the nucleic acid sequence of (a), (b), (c), or (d). In a further embodiment, nucleic acid sequences including a sequence encoding the active site domains or other functional motifs described above for several of the PUFA PKS domains are encompassed by the invention.

Particularly preferred embodiments of the present invention include isolated nucleic acid molecules encoding chimeric proteins useful in a PUFA PKS system as described herein. The present invention includes the use of any domain or protein from or derived from one PUFA PKS system in a domain and/or with proteins from or derived

from another PUFA PKS system in order to create novel PUFA PKS systems with unique qualities.

For example, one embodiment of the present invention relates to the use of a DH2 domain from a PUFA PKS system to modify a PUFA PKS system comprised of proteins/domains from a different organism or organisms, wherein the introduction of the DH2 domain (*e.g.,* in one embodiment, by substitution for the endogenous DH2 domain or similar domain in the host) modifies the ratio of PUFAs produced by the system, and particularly the ratio of omega-3 to omega-6 PUFAs produced by the system. This embodiment is described in detail below.

Some preferred nucleic acid molecules include a nucleic acid sequence encoding an amino acid sequence of SEQ ID NO:74, and biologically active fragments thereof, a nucleic acid sequence encoding an amino acid sequence that is at least about 60% identical to SEQ ID NO:74 having biological activity of at least one, two, three or more domains of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, or a nucleic acid sequence that is fully complementary to the nucleic acid sequences above. In one embodiment, the nucleic acid molecule includes a nucleic acid sequence selected from SEQ ID NO:73 and SEQ ID NO:75. In one embodiment, the nucleic acid molecule includes a nucleic acid sequence encoding the amino acid sequence encoded by a plasmid selected from the group of pDS49 and pDD24. In one embodiment, the nucleic acid molecule includes the nucleic acid sequence of a plasmid selected from the group of pDS49 and pDD24 that encodes a chimeric OrfC protein.

Other preferred embodiments including nucleic acid molecules comprising a nucleic acid sequence encoding a PUFA PKS protein or domain or homologue thereof from one PUFA PKS system, wherein the nucleic acid sequence is optimized for the codon usage of a different organism, such as a host in which the nucleic acid sequence is to be expressed. Examples of such nucleic acid sequences are described herein, and include, but are not limited to, the nucleic acid sequences represented by SEQ ID NO:70, SEQ ID NO:71, and SEQ ID NO:72, as well as SEQ ID NO:75. Codon optimized nucleic acid sequences encoding any PUFA PKS protein or domain, and particularly, any of the amino acid sequences described herein are encompassed by the invention. In one embodiment, such a nucleic acid molecule includes a nucleic acid sequence encoding the amino acid sequence encoded by a plasmid selected from the group of pThOrfC-synPS, pDD26, pDD32, or pDD24. In one embodiment, the nucleic acid molecule includes the

nucleic acid sequence of a plasmid selected from pThOrfC-synPS, pDD26, pDD32, or pDD24 that encodes a protein or chimeric protein useful in a PUFA PKS system.

According to the present invention, an amino acid sequence that has a biological activity of at least one domain of a PUFA PKS system is an amino acid sequence that has the biological activity of at least one domain of the PUFA PKS system described in detail herein, as exemplified by the *Schizochytrium* and *Thraustochytrium* PUFA PKS systems, and as further exemplified by the described biological activities of any of the proteins and domains in any of the PUFA PKS systems described in U.S. Patent No. 6,140,486, U.S. Patent No. 6,566,583, U.S. Patent Application Publication No. 20020194641, U.S. Patent Application Publication No. 20070089199, U.S. Patent Application Publication No. 20040235127, U.S. Patent Application Publication No. 20050100995, PCT Patent Publication No. WO 05/097982, or U.S. Patent Application Publication No. 20050014231, *supra*.

Accordingly, an isolated nucleic acid molecule of the present invention can encode the translation product of any PUFA PKS open reading frame, PUFA PKS domain, biologically active fragment thereof, or any homologue of a naturally occurring PUFA PKS open reading frame or domain which has biological activity. A homologue of given protein or domain is a protein or polypeptide that has an amino acid sequence which differs from the naturally occurring reference amino acid sequence (i.e., of the reference protein or domain) in that at least one or a few, but not limited to one or a few, amino acids have been deleted (e.g., a truncated version of the protein, such as a peptide or fragment), inserted, inverted, substituted and/or derivatized (e.g., by glycosylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, amidation and/or addition of glycosylphosphatidyl inositol). Preferred homologues of a PUFA PKS protein or domain are described in detail below. It is noted that homologues can include synthetically produced homologues, naturally occurring allelic variants of a given protein or domain, or homologous sequences from organisms other than the organism from which the reference sequence was derived.

In general, the biological activity or biological action of a protein or domain refers to any function(s) exhibited or performed by the protein or domain that is ascribed to the naturally occurring form of the protein or domain as measured or observed *in vivo* (i.e., in the natural physiological environment of the protein) or *in vitro* (i.e., under laboratory conditions). Biological activities of PUFA PKS systems and the individual

proteins/domains that make up a PUFA PKS system have been described in detail elsewhere herein. Modifications of a protein or domain, such as in a homologue or mimetic (discussed below), may result in proteins or domains having the same biological activity as the naturally occurring protein or domain, or in proteins or domains having decreased or increased biological activity as compared to the naturally occurring protein or domain. Modifications which result in a decrease in expression or a decrease in the activity of the protein or domain, can be referred to as inactivation (complete or partial), down-regulation, or decreased action of a protein or domain. Similarly, modifications which result in an increase in expression or an increase in the activity of the protein or domain, can be referred to as amplification, overproduction, activation, enhancement, up-regulation or increased action of a protein or domain. A functional domain of a PUFA PKS system is a domain (i.e., a domain can be a portion of a protein) that is capable of performing a biological function (i.e., has biological activity).

In accordance with the present invention, an isolated nucleic acid molecule is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. An isolated nucleic acid molecule can include a gene. An isolated nucleic acid molecule that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but typically no additional genes naturally found on the same chromosome, although some nucleic acid molecules may include nearby/linked genes that are not necessarily a part of the PUFA PKS gene or system. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecule can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Although the phrase "nucleic acid molecule" primarily refers to the physical nucleic acid molecule and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the nucleic acid molecule, the two

phrases can be used interchangeably, especially with respect to a nucleic acid molecule, or a nucleic acid sequence, being capable of encoding a protein or domain of a protein.

Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. Isolated nucleic acid molecules include natural nucleic acid molecules and homologues thereof, including, but not limited to, natural allelic variants and modified nucleic acid molecules in which nucleotides have been inserted, deleted, substituted, and/or inverted in such a manner that such modifications provide the desired effect on PUFA PKS system biological activity as described herein. Protein homologues (e.g., proteins encoded by nucleic acid homologues) have been discussed in detail above.

A nucleic acid molecule homologue can be produced using a number of methods known to those skilled in the art (see, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989). For example, nucleic acid molecules can be modified using a variety of techniques including, but not limited to, classic mutagenesis techniques and recombinant DNA techniques, such as site-directed mutagenesis, chemical treatment of a nucleic acid molecule to induce mutations, restriction enzyme cleavage of a nucleic acid fragment, ligation of nucleic acid fragments, PCR amplification and/or mutagenesis of selected regions of a nucleic acid sequence, synthesis of oligonucleotide mixtures and ligation of mixture groups to "build" a mixture of nucleic acid molecules and combinations thereof. Nucleic acid molecule homologues can be selected from a mixture of modified nucleic acids by screening for the function of the protein encoded by the nucleic acid and/or by hybridization with a wild-type gene.

The minimum size of a nucleic acid molecule of the present invention is a size sufficient to form a probe or oligonucleotide primer that is capable of forming a stable hybrid (e.g., under moderate, high or very high stringency conditions) with the complementary sequence of a nucleic acid molecule useful in the present invention, or of a size sufficient to encode an amino acid sequence having a biological activity of at least one domain of a PUFA PKS system according to the present invention. As such, the size of the nucleic acid molecule encoding such a protein can be dependent on nucleic acid composition and percent homology or identity between the nucleic acid molecule and complementary sequence as well as upon hybridization conditions *per se* (e.g., temperature, salt concentration, and formamide concentration). The minimal size of a

nucleic acid molecule that is used as an oligonucleotide primer or as a probe is typically at least about 12 to about 15 nucleotides in length if the nucleic acid molecules are GC-rich and at least about 15 to about 18 bases in length if they are AT-rich. There is no limit, other than a practical limit, on the maximal size of a nucleic acid molecule of the present invention, in that the nucleic acid molecule can include a sequence sufficient to encode a biologically active fragment of a domain of a PUFA PKS system, an entire domain of a PUFA PKS system, several domains within an open reading frame (Orf) of a PUFA PKS system, an entire Orf of a PUFA PKS system, or more than one Orf of a PUFA PKS system.

In one embodiment of the present invention, an isolated nucleic acid molecule comprises, consists essentially of, or consists of a nucleic acid sequence encoding an amino acid sequence selected from the group of: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO: 62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, or SEQ ID NO:74, or biologically active fragments thereof. In one aspect, the nucleic acid sequence is selected from: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, or SEQ ID NO:75.

In one embodiment of the present invention, any of the above-described PUFA PKS amino acid sequences, as well as homologues of such sequences, can be produced with from at least one, and up to about 20, additional heterologous amino acids flanking each of the C- and/or N-terminal end of the given amino acid sequence. The resulting protein or polypeptide can be referred to as "consisting essentially of" a given amino acid sequence. According to the present invention, the heterologous amino acids are a sequence of amino acids that are not naturally found (i.e., not found in nature, *in vivo*)

flanking the given amino acid sequence or which would not be encoded by the nucleotides that flank the naturally occurring nucleic acid sequence encoding the given amino acid sequence as it occurs in the gene, if such nucleotides in the naturally occurring sequence were translated using standard codon usage for the organism from which the given amino acid sequence is derived. Similarly, the phrase "consisting essentially of", when used with reference to a nucleic acid sequence herein, refers to a nucleic acid sequence encoding a given amino acid sequence that can be flanked by from at least one, and up to as many as about 60, additional heterologous nucleotides at each of the 5' and/or the 3' end of the nucleic acid sequence encoding the given amino acid sequence. The heterologous nucleotides are not naturally found (i.e., not found in nature, *in vivo*) flanking the nucleic acid sequence encoding the given amino acid sequence as it occurs in the natural gene.

The present invention also includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding an amino acid sequence having a biological activity of at least one domain of a PUFA PKS system. In one aspect, such a nucleic acid sequence encodes a homologue of any of the PUFA PKS proteins or domains described above, wherein the homologue has a biological activity of at least one (or two, three, four or more) domain of a PUFA PKS system as described previously herein.

In one aspect of the invention, a homologue of a PUFA PKS protein or domain encompassed by the present invention comprises an amino acid sequence that is at least about 60% identical to at least 500 consecutive amino acids of an amino acid sequence chosen from: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:39, SEQ ID NO:52, SEQ ID NO:62 or SEQ ID NO:74; wherein said amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the homologue is at least about 60% identical to at least about 600 consecutive amino acids, and more preferably to at least about 700 consecutive amino acids, and more preferably to at least about 800 consecutive amino acids, and more preferably to at least about 900 consecutive amino acids, and more preferably to at least about 1000 consecutive amino acids, and more preferably to at least about 1100 consecutive amino acids, and more preferably to at least about 1200 consecutive amino acids, and more preferably to at least about 1300 consecutive amino acids, and more preferably to at least about 1400 consecutive amino acids, and more preferably to at least about 1500 consecutive amino acids of any of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:39, SEQ ID NO:52, SEQ ID NO:62, or SEQ ID NO:74, or to the full

length of SEQ ID NO:6, SEQ ID NO:62, or SEQ ID NO:74. In a further aspect, the amino acid sequence of the homologue is at least about 60% identical to at least about 1600 consecutive amino acids, and more preferably to at least about 1700 consecutive amino acids, and more preferably to at least about 1800 consecutive amino acids, and more preferably to at least about 1900 consecutive amino acids, and more preferably to at least about 2000 consecutive amino acids of any of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:39, or SEQ ID NO:52 or to the full length of SEQ ID NO:4 or SEQ ID NO:52. In a further aspect, the amino acid sequence of the homologue is at least about 60% identical to at least about 2100 consecutive amino acids, and more preferably to at least about 2200 consecutive amino acids, and more preferably to at least about 2300 consecutive amino acids, and more preferably to at least about 2400 consecutive amino acids, and more preferably to at least about 2500 consecutive amino acids, and more preferably to at least about 2600 consecutive amino acids, and more preferably to at least about 2700 consecutive amino acids, and more preferably to at least about 2800 consecutive amino acids, and even more preferably, to the full length of SEQ ID NO:2 or SEQ ID NO:39.

In another aspect, a homologue of a PUFA PKS protein or domain encompassed by the present invention comprises an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to any of the above-described amino acid sequences, over any of the consecutive amino acid lengths described in the paragraphs above, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system.

In one aspect of the invention, a homologue of a PUFA PKS protein or domain encompassed by the present invention comprises an amino acid sequence that is at least about 60% identical to an amino acid sequence chosen from: SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:64, SEQ ID NO:66,

SEQ ID NO:68, or amino acid sequences comprising combinations of any of such amino acid sequences, wherein said amino acid sequence has a biological activity of at least one domain of a PUFA PKS system or accessory protein thereof. In a further aspect, the amino acid sequence of the homologue is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to any of the above-described amino acid sequences, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system or accessory protein thereof.

According to the present invention, the term "contiguous" or "consecutive", with regard to nucleic acid or amino acid sequences described herein, means to be connected in an unbroken sequence. For example, for a first sequence to comprise 30 contiguous (or consecutive) amino acids of a second sequence, means that the first sequence includes an unbroken sequence of 30 amino acid residues that is 100% identical to an unbroken sequence of 30 amino acid residues in the second sequence. Similarly, for a first sequence to have "100% identity" with a second sequence means that the first sequence exactly matches the second sequence with no gaps between nucleotides or amino acids.

As used herein, unless otherwise specified, reference to a percent (%) identity refers to an evaluation of homology which is performed using: (1) a BLAST 2.0 Basic BLAST homology search using blastp for amino acid searches, blastn for nucleic acid searches, and blastX for nucleic acid searches and searches of translated amino acids in all 6 open reading frames, all with standard default parameters, wherein the query sequence is filtered for low complexity regions by default (described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety); (2) a BLAST 2 alignment (using the parameters described below); (3) and/or PSI-BLAST with the standard default parameters (Position-Specific Iterated BLAST). It is noted that due to some differences in the standard parameters between BLAST 2.0 Basic BLAST and BLAST 2, two specific sequences might be recognized as having significant homology

using the BLAST 2 program, whereas a search performed in BLAST 2.0 Basic BLAST using one of the sequences as the query sequence may not identify the second sequence in the top matches. In addition, PSI-BLAST provides an automated, easy-to-use version of a "profile" search, which is a sensitive way to look for sequence homologues. The program first performs a gapped BLAST database search. The PSI-BLAST program uses the information from any significant alignments returned to construct a position-specific score matrix, which replaces the query sequence for the next round of database searching. Therefore, it is to be understood that percent identity can be determined by using any one of these programs.

Two specific sequences can be aligned to one another using BLAST 2 sequence as described in Tatusova and Madden, (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* 174:247-250, incorporated herein by reference in its entirety. BLAST 2 sequence alignment is performed in blastp or blastn using the BLAST 2.0 algorithm to perform a Gapped BLAST search (BLAST 2.0) between the two sequences allowing for the introduction of gaps (deletions and insertions) in the resulting alignment. For purposes of clarity herein, a BLAST 2 sequence alignment is performed using the standard default parameters as follows.

For blastn, using 0 BLOSUM62 matrix:
Reward for match = 1
Penalty for mismatch = -2
Open gap (5) and extension gap (2) penalties
gap x_dropoff (50) expect (10) word size (11) filter (on)

For blastp, using 0 BLOSUM62 matrix:
Open gap (11) and extension gap (1) penalties
gap x_dropoff (50) expect (10) word size (3) filter (on).

In another embodiment of the invention, an amino acid sequence having the biological activity of at least one domain of a PUFA PKS system of the present invention includes an amino acid sequence that is sufficiently similar to a naturally occurring PUFA PKS protein or polypeptide that a nucleic acid sequence encoding the amino acid sequence is capable of hybridizing under moderate, high, or very high stringency conditions (described below) to (i.e., with) a nucleic acid molecule encoding the naturally occurring PUFA PKS protein or polypeptide (i.e., to the complement of the nucleic acid strand encoding the naturally occurring PUFA PKS protein or polypeptide). Preferably, an amino acid sequence having the biological activity of at least one domain of a PUFA PKS

system of the present invention is encoded by a nucleic acid sequence that hybridizes under moderate, high or very high stringency conditions to the complement of a nucleic acid sequence that encodes a protein comprising an amino acid sequence represented by any of the amino acid sequences described herein.

In another embodiment of the invention, a nucleotide sequence of the present invention is a nucleotide sequence isolated from (obtainable from), identical to, or a homologue of, the nucleotide sequence from a *Schizochytrium,* wherein the nucleotide sequence from a *Schizochytrium* (including either strand of a DNA molecule from *Schizochytrium*) hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence encoding an amino acid sequence represented by any of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:32. In one embodiment, the *Schizochytrium* is *Schizochytrium* ATCC 20888. In another embodiment, the *Schizochytrium* is a daughter strain of *Schizochytrium* 20888, including mutated strains thereof (*e.g.,* N230D). In one embodiment, the nucleic acid sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence selected from: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:31.

In another embodiment of the invention, a nucleotide sequence of the present invention is a nucleotide sequence isolated from (obtainable from), identical to, or a homologue of, the nucleotide sequence from a *Thraustochytrium,* wherein the nucleotide sequence from a *Thraustochytrium* (including either strand of a DNA molecule from *Thraustochytrium*) hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence encoding an amino acid sequence represented by any of SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68. In one embodiment, the *Thraustochytrium* is *Thraustochytrium* 23B (ATCC 20892). In one embodiment, the nucleic acid sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence selected from: SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ

ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, or SEQ ID NO:67.

In yet another embodiment, a nucleotide sequence of the present invention is a nucleotide sequence isolated from (obtainable from), identical to, or a homologue of, the nucleotide sequence from a eukaryotic organism (*e.g.,* a thraustochytrid or a labyrinthulid) or a marine bacterium, wherein the nucleotide sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence encoding any of the amino acid sequences represented herein.

In another embodiment, a nucleotide sequence of the present invention is a nucleotide sequence isolated from (obtainable from), identical to, or a homologue of, any nucleotide sequence encoding an accessory protein described herein (including either strand of a DNA molecule), where, in one embodiment, the nucleotide sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence encoding an amino acid sequence represented by SEQ ID NO:34. In one embodiment, the nucleic acid sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence represented by SEQ ID NO:33.

In another embodiment, a nucleotide sequence of the present invention is a nucleotide sequence isolated from (obtainable from), identical to, or a homologue of, any codon-optimized or chimeric nucleotide sequence described herein (including either strand of a DNA molecule), where, in one embodiment, the nucleotide sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence encoding an amino acid sequence represented by SEQ ID NO:74. In one embodiment, the nucleic acid sequence hybridizes under moderate, high, or very high stringency conditions to a nucleotide sequence selected from SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, or SEQ ID NO:75.

Methods to deduce a complementary sequence are known to those skilled in the art. It should be noted that since amino acid sequencing and nucleic acid sequencing technologies are not entirely error-free, the sequences presented herein, at best, represent apparent sequences of PUFA PKS domains and proteins of the present invention, or of the nucleotide sequences encoding such amino acid sequences.

As used herein, hybridization conditions refer to standard hybridization conditions under which nucleic acid molecules are used to identify similar nucleic acid molecules.

Such standard conditions are disclosed, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety (see specifically, pages 9.31-9.62). In addition, formulae to calculate the appropriate hybridization and wash conditions to achieve hybridization permitting varying degrees of mismatch of nucleotides are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284; Meinkoth et al., *ibid.*, is incorporated by reference herein in its entirety.

More particularly, moderate stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 70% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 30% or less mismatch of nucleotides). High stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 80% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 20% or less mismatch of nucleotides). Very high stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 90% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 10% or less mismatch of nucleotides). As discussed above, one of skill in the art can use the formulae in Meinkoth et al., *ibid.* to calculate the appropriate hybridization and wash conditions to achieve these particular levels of nucleotide mismatch. Such conditions will vary, depending on whether DNA:RNA or DNA:DNA hybrids are being formed. Calculated melting temperatures for DNA:DNA hybrids are 10°C less than for DNA:RNA hybrids. In particular embodiments, stringent hybridization conditions for DNA:DNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 20°C and about 35°C (lower stringency), more preferably, between about 28°C and about 40°C (more stringent), and even more preferably, between about 35°C and about 45°C (even more stringent), with appropriate wash conditions. In particular embodiments, stringent hybridization conditions for DNA:RNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 30°C and about 45°C, more preferably, between about 38°C and about 50°C, and even more preferably, between about 45°C and about 55°C, with similarly stringent wash

conditions. These values are based on calculations of a melting temperature for molecules larger than about 100 nucleotides, 0% formamide and a G + C content of about 40%. Alternatively, T_m can be calculated empirically as set forth in Sambrook et al., *supra*, pages 9.31 to 9.62. In general, the wash conditions should be as stringent as possible, and should be appropriate for the chosen hybridization conditions. For example, hybridization conditions can include a combination of salt and temperature conditions that are approximately 20-25°C below the calculated T_m of a particular hybrid, and wash conditions typically include a combination of salt and temperature conditions that are approximately 12-20°C below the calculated T_m of the particular hybrid. One example of hybridization conditions suitable for use with DNA:DNA hybrids includes a 2-24 hour hybridization in 6X SSC (50% formamide) at about 42°C, followed by washing steps that include one or more washes at room temperature in about 2X SSC, followed by additional washes at higher temperatures and lower ionic strength (e.g., at least one wash as about 37°C in about 0.1X-0.5X SSC, followed by at least one wash at about 68°C in about 0.1X-0.5X SSC).

Yet another embodiment of the present invention includes a nucleic acid molecule comprising, consisting essentially of, or consisting of, a nucleic acid sequence that is identical to, or that is a homologue of (as defined above) the nucleic acid sequence of a plasmid selected from: pJK1126 (ATCC Accession No. PTA-7648), pJK1129 (ATCC Accession No. PTA-7649), pJK1131 (ATCC Accession No. PTA-7650), pJK306 (ATCC Accession No. PTA-7641), pJK320 (ATCC Accession No. PTA-7644), pJK324 (ATCC Accession No. PTA-7643), pBR002 (ATCC Accession No. PTA-7642), Th23BOrfA_pBR812.1 (ATCC Accession No. PTA-8232) Th23BOrfA_pBR811 (ATCC Accession No. PTA-8231), Th23BOrfB_pBR800 (ATCC Accession No. PTA-8227) or Th23BOrfC_pBR709A (ATCC Accession No. PTA-8228).

In another embodiment, the present invention includes a nucleic acid molecule comprising, consisting essentially of, or consisting of, a nucleic acid sequence that is identical to, or that is a homologue of (as defined above), the nucleic acid sequence of a plasmid selected from: pThOrfC-synPS (ATCC Accession No. PTA-8229), pDS49 (ATCC Accession No. PTA-8230), pDD24 (ATCC Accession No. PTA-8226), pDD26 (ATCC Accession No. PTA-8411), pDD32 (ATCC Accession No. PTA-8412), or OrfB*_pJK780 (ATCC Accession No. PTA-8225).

Yet another embodiment of the present invention includes a nucleic acid molecule comprising, consisting essentially of, or consisting of, a nucleic acid sequence that encodes an amino acid sequence that is identical to, or that is a homologue of (as defined above) the amino acid sequence encoded by a plasmid selected from: pJK1126 (ATCC Accession No. PTA-7648), pJK1129 (ATCC Accession No. PTA-7649), pJK1131 (ATCC Accession No. PTA-7650), pJK306 (ATCC Accession No. PTA-7641), pJK320 (ATCC Accession No. PTA-7644), pJK324 (ATCC Accession No. PTA-7643), pBR002 (ATCC Accession No. PTA-7642), Th23BOrfA_pBR812.1 (ATCC Accession No. PTA-8232) Th23BOrfA_pBR811 (ATCC Accession No. PTA-8231), Th23BOrfB_pBR800 (ATCC Accession No. PTA-8227) or Th23BOrfC_pBR709A (ATCC Accession No. PTA-8228).

In another embodiment, the present invention includes a nucleic acid molecule comprising, consisting essentially of, or consisting of, a nucleic acid sequence that encodes an amino acid sequence that is identical to, or that is a homologue of (as defined above) the amino acid sequence encoded by a plasmid selected from: pThOrfC-synPS (ATCC Accession No. PTA-8229), pDS49 (ATCC Accession No. PTA-8230), pDD24 (ATCC Accession No. PTA-8226), pDD26 (ATCC Accession No. PTA-8411), pDD32 (ATCC Accession No. PTA-8412), or OrfB*_pJK780 (ATCC Accession No. PTA-8225).

Another embodiment of the present invention includes a recombinant nucleic acid molecule comprising a recombinant vector and a nucleic acid molecule comprising a nucleic acid sequence encoding an amino acid sequence having a biological activity of at least one domain or protein of a PUFA PKS system as described herein. Such nucleic acid sequences and domains or proteins are described in detail above. According to the present invention, a recombinant vector is an engineered (i.e., artificially produced) nucleic acid molecule that is used as a tool for manipulating a nucleic acid sequence of choice and for introducing such a nucleic acid sequence into a host cell. The recombinant vector is therefore suitable for use in cloning, sequencing, and/or otherwise manipulating the nucleic acid sequence of choice, such as by expressing and/or delivering the nucleic acid sequence of choice into a host cell to form a recombinant cell. Such a vector typically contains heterologous nucleic acid sequences, that is nucleic acid sequences that are not naturally found adjacent to nucleic acid sequence to be cloned or delivered, although the vector can also contain regulatory nucleic acid sequences (e.g., promoters, untranslated regions) which are naturally found adjacent to nucleic acid molecules of the present invention or which are useful for expression of the nucleic acid molecules of the present

invention (discussed in detail below). The vector can be either RNA or DNA, either prokaryotic or eukaryotic, and typically is a plasmid. The vector can be maintained as an extrachromosomal element (e.g., a plasmid) or it can be integrated into the chromosome of a recombinant organism (e.g., a microbe or a plant). The entire vector can remain in place within a host cell, or under certain conditions, the plasmid DNA can be deleted, leaving behind the nucleic acid molecule of the present invention. The integrated nucleic acid molecule can be under chromosomal promoter control, under native or plasmid promoter control, or under a combination of several promoter controls. Single or multiple copies of the nucleic acid molecule can be integrated into the chromosome. A recombinant vector of the present invention can contain at least one selectable marker.

In one embodiment, a recombinant vector used in a recombinant nucleic acid molecule of the present invention is an expression vector. As used herein, the phrase "expression vector" is used to refer to a vector that is suitable for production of an encoded product (e.g., a protein of interest). In this embodiment, a nucleic acid sequence encoding the product to be produced (e.g., a PUFA PKS domain) is inserted into the recombinant vector to produce a recombinant nucleic acid molecule. The nucleic acid sequence encoding the protein to be produced is inserted into the vector in a manner that operatively links the nucleic acid sequence to regulatory sequences in the vector which enable the transcription and translation of the nucleic acid sequence within the recombinant host cell.

In another embodiment, a recombinant vector used in a recombinant nucleic acid molecule of the present invention is a targeting vector. As used herein, the phrase "targeting vector" is used to refer to a vector that is used to deliver a particular nucleic acid molecule into a recombinant host cell, wherein the nucleic acid molecule is used to delete or inactivate an endogenous gene within the host cell or microorganism (i.e., used for targeted gene disruption or knock-out technology). Such a vector may also be known in the art as a "knock-out" vector. In one aspect of this embodiment, a portion of the vector, but more typically, the nucleic acid molecule inserted into the vector (i.e., the insert), has a nucleic acid sequence that is homologous to a nucleic acid sequence of a target gene in the host cell (i.e., a gene which is targeted to be deleted or inactivated). The nucleic acid sequence of the vector insert is designed to bind to the target gene such that the target gene and the insert undergo homologous recombination, whereby the endogenous target gene is deleted, inactivated or attenuated (i.e., by at least a portion of the endogenous target gene being mutated or deleted).

Typically, a recombinant nucleic acid molecule includes at least one nucleic acid molecule of the present invention operatively linked to one or more transcription control sequences. As used herein, the phrase "recombinant molecule" or "recombinant nucleic acid molecule" primarily refers to a nucleic acid molecule or nucleic acid sequence operatively linked to a transcription control sequence, but can be used interchangeably with the phrase "nucleic acid molecule", when such nucleic acid molecule is a recombinant molecule as discussed herein. According to the present invention, the phrase "operatively linked" refers to linking a nucleic acid molecule to a transcription control sequence in a manner such that the molecule is able to be expressed when transfected (i.e., transformed, transduced, transfected, conjugated or conduced) into a host cell. Transcription control sequences are sequences which control the initiation, elongation, or termination of transcription. Particularly important transcription control sequences are those which control transcription initiation, such as promoter, enhancer, operator and repressor sequences. Suitable transcription control sequences include any transcription control sequence that can function in a host cell or organism into which the recombinant nucleic acid molecule is to be introduced.

Recombinant nucleic acid molecules of the present invention can also contain additional regulatory sequences, such as translation regulatory sequences, origins of replication, and other regulatory sequences that are compatible with the recombinant cell. In one embodiment, a recombinant molecule of the present invention, including those which are integrated into the host cell chromosome, also contains secretory signals (i.e., signal segment nucleic acid sequences) to enable an expressed protein to be secreted from the cell that produces the protein. Suitable signal segments include a signal segment that is naturally associated with the protein to be expressed or any heterologous signal segment capable of directing the secretion of the protein according to the present invention. In another embodiment, a recombinant molecule of the present invention comprises a leader sequence to enable an expressed protein to be delivered to and inserted into the membrane of a host cell. Suitable leader sequences include a leader sequence that is naturally associated with the protein, or any heterologous leader sequence capable of directing the delivery and insertion of the protein to the membrane of a cell.

The present inventors have found that the *Schizochytrium* and *Thraustochytrium* PUFA PKS Orfs A and B are closely linked in the genome and the region between the Orfs has been sequenced. In *Schizochytrium,* the Orfs are oriented in opposite directions

and 4244 base pairs separate the start (ATG) codons (i.e. they are arranged as follows: 3'OrfA5'-4244 bp-5'OrfB3'). Examination of the 4244 bp intergenic region did not reveal any obvious Orfs (no significant matches were found on a BlastX search). Both Orfs A and B are highly expressed in *Schizochytrium*, at least during the time of oil production, implying that active promoter elements are embedded in this intergenic region. These genetic elements are believed to have utility as a bi-directional promoter sequence for transgenic applications. For example, in a preferred embodiment, one could clone this region, place any genes of interest at each end and introduce the construct into *Schizochytrium* (or some other host in which the promoters can be shown to function). It is predicted that the regulatory elements, under the appropriate conditions, would provide for coordinated, high level expression of the two introduced genes. The complete nucleotide sequence for the regulatory region containing *Schizochytrium* PUFA PKS regulatory elements (e.g., a promoter) is represented herein as SEQ ID NO:76.

In a similar manner, OrfC is highly expressed in *Schizochytrium* during the time of oil production and regulatory elements are expected to reside in the region upstream of its start codon. A region of genomic DNA upstream of OrfC has been cloned and sequenced and is represented herein as (SEQ ID NO:77). This sequence contains the 3886 nt immediately upstream of the OrfC start codon. Examination of this region did not reveal any obvious Orfs (i.e., no significant matches were found on a BlastX search). It is believed that regulatory elements contained in this region, under the appropriate conditions, will provide for high-level expression of a gene placed behind them. Additionally, under the appropriate conditions, the level of expression may be coordinated with genes under control of the A-B intergenic region (SEQ ID NO:76).

Therefore, in one embodiment, a recombinant nucleic acid molecule useful in the present invention, as disclosed herein, can include a PUFA PKS regulatory region contained within SEQ ID NO:76 and/or SEQ ID NO:77. Such a regulatory region can include any portion (fragment) of SEQ ID NO:76 and/or SEQ ID NO:77 that has at least basal PUFA PKS transcriptional activity (at least basal promoter activity).

One or more recombinant molecules of the present invention can be used to produce an encoded product (e.g., a PUFA PKS domain, protein, or system) of the present invention. In one embodiment, an encoded product is produced by expressing a nucleic acid molecule as described herein under conditions effective to produce the protein. A preferred method to produce an encoded protein is by transfecting a host cell with one or

more recombinant molecules to form a recombinant cell. Suitable host cells to transfect include, but are not limited to, any bacterial, fungal (e.g., yeast), insect, plant or animal cell that can be transfected. Host cells can be either untransfected cells or cells that are already transfected with at least one other recombinant nucleic acid molecule.

According to the present invention, the term "transfection" is used to refer to any method by which an exogenous nucleic acid molecule (i.e., a recombinant nucleic acid molecule) can be inserted into a cell. The term "transformation" can be used interchangeably with the term "transfection" when such term is used to refer to the introduction of nucleic acid molecules into microbial cells, such as algae, bacteria and yeast. In microbial systems, the term "transformation" is used to describe an inherited change due to the acquisition of exogenous nucleic acids by the microorganism and is essentially synonymous with the term "transfection." However, in animal cells, transformation has acquired a second meaning which can refer to changes in the growth properties of cells in culture after they become cancerous, for example. Therefore, to avoid confusion, the term "transfection" is preferably used with regard to the introduction of exogenous nucleic acids into animal cells, and the term "transfection" will be used herein to generally encompass transfection of animal cells, plant cells and transformation of microbial cells, to the extent that the terms pertain to the introduction of exogenous nucleic acids into a cell. Therefore, transfection techniques include, but are not limited to, transformation, particle bombardment, electroporation, microinjection, lipofection, adsorption, infection and protoplast fusion.

It will be appreciated by one skilled in the art that use of recombinant DNA technologies can improve control of expression of transfected nucleic acid molecules by manipulating, for example, the number of copies of the nucleic acid molecules within the host cell, the efficiency with which those nucleic acid molecules are transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Additionally, the promoter sequence might be genetically engineered to improve the level of expression as compared to the native promoter. Recombinant techniques useful for controlling the expression of nucleic acid molecules include, but are not limited to, integration of the nucleic acid molecules into one or more host cell chromosomes, addition of vector stability sequences to plasmids, substitutions or modifications of transcription control signals (e.g., promoters, operators, enhancers), substitutions or modifications of translational control signals (e.g., ribosome binding sites,

Shine-Dalgarno sequences), modification of nucleic acid molecules to correspond to the codon usage of the host cell, and deletion of sequences that destabilize transcripts.

General discussion above with regard to recombinant nucleic acid molecules and transfection of host cells is intended to be applied to any recombinant nucleic acid molecule discussed herein, including those encoding any amino acid sequence having a biological activity of at least one domain from a PUFA PKS, those encoding amino acid sequences from other PKS systems, and those encoding other proteins or domains.

This invention also relates to PUFA PKS systems (and proteins or domains thereof) from microorganisms other than those described specifically herein that are homologous in structure, domain organization and/or function to any of the PUFA PKS system (and proteins or domains thereof) as described herein. In addition, this invention relates to use of these microorganisms and the PUFA PKS systems or components thereof (*e.g.,* DH2 domains) from these microorganisms in the various applications for a PUFA PKS system (e.g., genetically modified organisms and methods of producing bioactive molecules) according to the present invention. A screening process for identification of microorganisms comprising a PUFA PKS system is described in detail in U.S. Patent Application Publication No. 20020194641, *supra*. The knowledge of the structure and function of the PUFA PKS proteins and domains described herein, and the nucleotide sequence encoding the same, are useful tools for the identification, confirmation, and/or isolation of homologues of such proteins or polynucleotides.

According to the present invention, the term "thraustochytrid" refers to any members of the order Thraustochytriales, which includes the family Thraustochytriaceae, and the term "labyrinthulid" refers to any member of the order Labyrinthulales, which includes the family Labyrinthulaceae. The members of the family Labyrinthulaceae were at one time considered to be members of the order Thraustochytriales, but in more recent revisions of the taxonomy of such organisms, the family is now considered to be a member of the order Labyrinthulales, and both Labyrinthulales and Thraustochytriales are considered to be members of the phylum Labyrinthulomycota. Developments have resulted in frequent revision of the taxonomy of the thraustochytrids and labyrinthulids. However, taxonomic theorists now generally place both of these groups of microorganisms with the algae or algae-like protists within the Stramenopile lineage. The current taxonomic placement of the thraustochytrids and labyrinthulids can be summarized as follows:

Realm: Stramenopila (Chromista)
Phylum: Labyrinthulomycota
Class: Labyrinthulomycetes
Order: Labyrinthulales
Family: Labyrinthulaceae
Order: Thraustochytriales
Family: Thraustochytriaceae

However, because of remaining taxonomic uncertainties it would be best for the purposes of the present invention to consider the strains described in the present invention as thraustochytrids to include the following organisms: Order: Thraustochytriales; Family: Thraustochytriaceae; Genera: *Thraustochytrium* (Species: sp., *arudimentale, aureum, benthicola, globosum, kinnei, motivum, multirudimentale, pachydermum, proliferum, roseum, striatum*), *Ulkenia* (Species: sp., *amoeboidea, kerguelensis, minuta, profunda, radiata, sailens, sarkariana, schizochytrops, visurgensis, yorkensis*), *Schizochytrium* (Species: sp., *aggregatum, limnaceum, mangrovei, minutum, octosporum*), *Japonochytrium* (Species: sp., *marinum*), *Aplanochytrium* (Species: sp., *haliotidis, kerguelensis, profunda, stocchinoi*), *Althornia* (Species: sp., *crouchii*), or *Elina* (Species: sp., *marisalba, sinorifica*). It is to be noted that the original description of the genus *Ulkenia* was not published in a peer-reviewed journal so some questions remain as to the validity of this genus and the species placed within it. For the purposes of this invention, species described within *Ulkenia* will be considered to be members of the genus *Thraustochytrium*.

Strains described in the present invention as Labyrinthulids include the following organisms: Order: Labyrinthulales, Family:Labyrinthulaceae, Genera: *Labyrinthula* (Species: sp., *algeriensis, coenocystis, chattonii, macrocystis, macrocystis atlantica, macrocystis macrocystis, marina, minuta, roscoffensis, valkanovii, vitellina, vitellina pacifica, vitellina vitellina, zopfii*), *Labyrinthuloides* (Species: sp., *haliotidis*, *yorkensis*), *Labyrinthomyxa* (Species: sp., *marina*), *Diplophrys* (Species: sp., *archeri*), *Pyrrhosorus* (Species: sp., *marinus*), *Sorodiplophrys* (Species: sp., *stercorea*) or *Chlamydomyxa* (Species: sp., *labyrinthuloides*, *montana*) (although there is currently not a consensus on the exact taxonomic placement of *Pyrrhosorus, Sorodiplophrys* or *Chlamydomyxa*).

To produce significantly high yields of various bioactive molecules using the PUFA PKS system of the present invention, an organism, preferably a microorganism or a

plant or plant part (*e.g.,* a plant cell), can be genetically modified to affect the activity of a PUFA PKS system. In one aspect, such an organism can endogenously contain and express a PUFA PKS system, and the genetic modification can be a genetic modification of one or more of the functional domains of the endogenous PUFA PKS system, whereby the modification has some effect on the activity of the PUFA PKS system. In another aspect, such an organism can endogenously contain and express a PUFA PKS system, and the genetic modification can be an introduction of at least one exogenous nucleic acid sequence (e.g., a recombinant nucleic acid molecule), wherein the exogenous nucleic acid sequence encodes at least one biologically active domain or protein from the same or a second PKS system and/or a protein that affects the activity of said PUFA PKS system (e.g., a phosphopantetheinyl transferases (PPTase), discussed below). In yet another aspect, the organism does not necessarily endogenously (naturally) contain a PUFA PKS system, but is genetically modified to introduce at least one recombinant nucleic acid molecule encoding an amino acid sequence having the biological activity of at least one domain of a PUFA PKS system. In this aspect, PUFA PKS activity is affected by introducing or increasing PUFA PKS activity in the organism. Various embodiments associated with each of these aspects will be discussed in greater detail below.

Therefore, according to the present invention, one embodiment relates to a genetically modified microorganism, wherein the microorganism expresses a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. The at least one domain of the PUFA PKS system is encoded by a nucleic acid sequence described herein. The genetic modification affects the activity of the PKS system in the organism. The genetically modified microorganism can include any one or more of the above-identified nucleic acid sequences, and/or any of the other homologues of any of the PUFA PKS ORFs or domains as described in detail above.

As used herein, a genetically modified microorganism can include a genetically modified bacterium, protist, microalgae, fungus, or other microbe, and particularly, any of the genera of the order Thraustochytriales (e.g., a thraustochytrid) described herein. Such a genetically modified microorganism has a genome which is modified (i.e., mutated or changed) from its normal (i.e., wild-type or naturally occurring) form such that the desired result is achieved (i.e., increased or modified PUFA PKS activity and/or production of a desired product using the PUFA PKS system or component thereof). Genetic modification of a microorganism can be accomplished using classical strain development and/or

molecular genetic techniques. Such techniques known in the art and are generally disclosed for microorganisms, for example, in Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press. The reference Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety. A genetically modified microorganism can include a microorganism in which nucleic acid molecules have been inserted, deleted or modified (i.e., mutated; e.g., by insertion, deletion, substitution, and/or inversion of nucleotides), in such a manner that such modifications provide the desired effect within the microorganism.

Preferred microorganism host cells to modify according to the present invention include, but are not limited to, any bacteria, protist, microalga, fungus, or protozoa. In one aspect, preferred microorganisms to genetically modify include, but are not limited to, any microorganism of the order Thraustochytriales or any microorganism of the order Labyrinthulales. Particularly preferred host cells for use in the present invention could include microorganisms from a genus including, but not limited to: *Thraustochytrium, Ulkenia, Schizochytrium, Japonochytrium, Aplanochytrium, Althornia, Elina, Labyrinthula, Labyrinthuloides, Labyrinthomyxa, Diplophrys, Pyrrhosorus, Sorodiplophrys* or *Chlamydomyxa.* Other examples of suitable host microorganisms for genetic modification include, but are not limited to, yeast including *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, or other yeast such as *Candida, Kluyveromyces*, or other fungi, for example, filamentous fungi such as *Aspergillus, Neurospora, Penicillium*, etc. Bacterial cells also may be used as hosts. This includes *Escherichia coli*, which can be useful in fermentation processes. Alternatively, a host such as a *Lactobacillus* species or *Bacillus* species can be used as a host.

Another embodiment of the present invention relates to a genetically modified plant or part of a plant (*e.g.,* wherein the plant has been genetically modified to express a PUFA PKS system described herein), which includes at least the core PUFA PKS enzyme complex and, in one embodiment, at least one PUFA PKS accessory protein, (*e.g.,* a PPTase), so that the plant produces PUFAs. Preferably, the plant is an oil seed plant, wherein the oil seeds or oil in the oil seeds contain PUFAs produced by the PUFA PKS system. Such oils contain a detectable amount of at least one target or primary PUFA that is the product of the PUFA PKS system. Plants are not known to endogenously contain a PUFA PKS system, and therefore, the PUFA PKS systems of the present invention represent an opportunity to produce plants with unique fatty acid production capabilities.

It is a particularly preferred embodiment of the present invention to genetically engineer plants to produce one or more PUFAs in the same plant, including, EPA, DHA, DPA (n-3 and/or n-6), ARA, GLA, SDA and others. The present invention offers the ability to create any one of a number of "designer oils" in various ratios and forms.

Methods for the genetic engineering of plants are well known in the art. For instance, numerous methods for plant transformation have been developed, including biological and physical transformation protocols. See, for example, Miki et al., "Procedures for Introducing Foreign DNA into Plants" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pp. 67-88. In addition, vectors and *in vitro* culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber et al., "Vectors for Plant Transformation" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pp. 89-119.

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. See, for example, Horsch et al., *Science* **227**:1229 (1985). *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. See, for example, Kado, C.I., *Crit. Rev. Plant. Sci.* **10**:1 (1991). Descriptions of *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided by numerous references, including Gruber et al., supra, Miki et al., supra, Moloney et al., *Plant Cell Reports* **8**:238 (1989), and U.S. Patents Nos. 4,940,838 and 5,464,763.

Another generally applicable method of plant transformation is microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds sufficient to penetrate plant cell walls and membranes. Sanford et al., *Part. Sci. Technol.* **5**:27 (1987), Sanford, J.C., *Trends Biotech.* **6**:299 (1988), Sanford, J.C., *Physiol. Plant* **79**:206 (1990), Klein et al., *Biotechnology* **10**:268 (1992).

Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., *Bio/Technology* **9**:996 (1991). Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants. Deshayes et al., *EMBO*

J., **4**:2731 (1985), Christou et al., *Proc Natl. Acad. Sci. USA* **84**:3962 (1987). Direct uptake of DNA into protoplasts using $CaCl_2$ precipitation, polyvinyl alcohol or poly-L-ornithine have also been reported. Hain et al., *Mol. Gen. Genet.* **199**:161 (1985) and Draper et al., *Plant Cell Physiol.* **23**:451 (1982). Electroporation of protoplasts and whole cells and tissues have also been described. Donn et al., In Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, p. 53 (1990); D'Halluin et al., *Plant Cell* **4**:1495-1505 (1992) and Spencer et al., *Plant Mol. Biol.* **24**:51-61 (1994).

Following the introduction of the genetic construct into plant cells, plant cells are grown and upon emergence of differentiating tissue such as shoots and roots, mature plants are generated. Typically a plurality of plants is generated. Methodologies for regenerating plants will be generally known to those skilled in the art and may be found in for example: Plant Cell and Tissue Culture, 1994, Vasil and Thorpe Eds. Kluwer Academic Publishers and in: Plant Cell Culture Protocols (Methods in Molecular Biology 111, 1999 Hall Eds Humana Press).

As used herein, a genetically modified plant can include any genetically modified plant including higher plants and particularly, any consumable plants or plants useful for producing a desired bioactive molecule of the present invention. "Plant parts", as used herein, include any parts of a plant, including, but not limited to, seeds (immature or mature), oils, pollen, embryos, flowers, fruits, shoots, leaves, roots, stems, explants, etc. A genetically modified plant has a genome that is modified (i.e., mutated or changed) from its normal (i.e., wild-type or naturally occurring) form such that the desired result is achieved (e.g., PUFA PKS activity and production of PUFAs). Genetic modification of a plant can be accomplished using classical strain development and/or molecular genetic techniques. Methods for producing a transgenic plant, wherein a recombinant nucleic acid molecule encoding a desired amino acid sequence is incorporated into the genome of the plant, are known in the art. A preferred plant to genetically modify according to the present invention is preferably a plant suitable for consumption by animals, including humans.

Preferred plants to genetically modify according to the present invention (*i.e.,* plant host cells) include, but are not limited to any higher plants, including both dicotyledonous and monocotyledonous plants, and particularly consumable plants, including crop plants and especially plants used for their oils. Such plants can include, but are not limited to, for example: canola, soybeans, rapeseed, linseed, corn, safflowers, sunflowers and tobacco.

Thus, any plant species or plant cell may be selected. Particular cells used herein, and plants grown or derived therefrom, include, but are not limited to, cells obtainable from canola (*Brassica rapa L.*); soybean (*Glycine max*); rapeseed (*Brassica spp.*); linseed/flax (*Linum usitatissimum*); maize (corn) (*Zea mays*); safflower (*Carthamus tinctorius*); sunflower (*Helianthus annuus*); tobacco (*Nicotiana tabacum*); *Arabidopsis thaliana*, Brazil nut (*Betholettia excelsa*); castor bean (*Riccinus communis*); coconut (*Cocus nucifera*); coriander (*Coriandrum sativum*); cotton (*Gossypium spp.*); groundnut (*Arachis hypogaea*); jojoba (*Simmondsia chinensis*); mustard (*Brassica spp.* and *Sinapis alba*); oil palm (*Elaeis guineeis*); olive (*Olea eurpaea*); rice (*Oryza sativa*); squash (*Cucurbita maxima*); barley (*Hordeum vulgare*); wheat (*Traeticum aestivum*); and duckweed (*Lemnaceae sp.*). It should be noted that in accordance herewith the genetic background within a plant species may vary.

Other preferred plants include those plants that are known to produce compounds used as pharmaceutical agents, flavoring agents, nutraceutical agents, functional food ingredients or cosmetically active agents or plants that are genetically engineered to produce these compounds/agents.

In a further embodiment plant cell cultures may be used in accordance herewith. In such embodiments plant cells are not grown into differentiated plants and cultivated using ordinary agricultural practices, but instead grown and maintained in a liquid medium.

According to the present invention, a genetically modified microorganism or plant includes a microorganism or plant that has been modified using recombinant technology. As used herein, genetic modifications that result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e., the protein encoded by the gene) can be referred to as inactivation (complete or partial), deletion, interruption, blockage or down-regulation of a gene. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such gene, can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein does not exist), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no enzymatic activity or action). Genetic modifications that result in an increase in gene expression or function can be referred to as amplification,

overproduction, overexpression, activation, enhancement, addition, or up-regulation of a gene.

The genetic modification of a microorganism or plant according to the present invention preferably affects the activity of the PKS system expressed by the plant, whether the PKS system is endogenous and genetically modified, endogenous with the introduction of recombinant nucleic acid molecules into the organism, or provided completely by recombinant technology. According to the present invention, to "affect the activity of a PKS system" includes any genetic modification that causes any detectable or measurable change or modification in the PKS system expressed by the organism as compared to in the absence of the genetic modification. A detectable change or modification in the PKS system can include, but is not limited to: the introduction of PKS system activity into an organism such that the organism now has measurable/detectable PKS system activity (i.e., the organism did not contain a PKS system prior to the genetic modification), the introduction into the organism of a functional domain from a different PKS system than a PKS system endogenously expressed by the organism such that the PKS system activity is modified (e.g., DH2 domain from one PUFA PKS system is introduced into the PUFA PKS system of an different organism), a change in the amount of a bioactive molecule produced by the PKS system (e.g., the system produces more (increased amount) or less (decreased amount) of a given product as compared to in the absence of the genetic modification), a change in the type of a bioactive molecule produced by the PKS system (e.g., the system produces a new or different product, or a variant of a product that is naturally produced by the system), and/or a change in the ratio of multiple bioactive molecules produced by the PKS system (e.g., the system produces a different ratio of one PUFA to another PUFA, produces a completely different lipid profile as compared to in the absence of the genetic modification, or places various PUFAs in different positions in a triacylglycerol as compared to the natural configuration). Such a genetic modification includes any type of genetic modification and specifically includes modifications made by recombinant technology and by classical mutagenesis.

It should be noted that reference to increasing the activity of a functional domain or protein in a PUFA PKS system refers to any genetic modification in the organism containing the domain or protein (or into which the domain or protein is to be introduced) which results in increased functionality of the domain or protein system and can include higher activity of the domain or protein (e.g., specific activity or *in vivo* enzymatic

activity), reduced inhibition or degradation of the domain or protein system, and overexpression of the domain or protein. For example, gene copy number can be increased, expression levels can be increased by use of a promoter that gives higher levels of expression than that of the native promoter, or a gene can be altered by genetic engineering or classical mutagenesis to increase the activity of the domain or protein encoded by the gene.

Similarly, reference to decreasing the activity of a functional domain or protein in a PUFA PKS system refers to any genetic modification in the organism containing such domain or protein (or into which the domain or protein is to be introduced) which results in decreased functionality of the domain or protein and includes decreased activity of the domain or protein, increased inhibition or degradation of the domain or protein and a reduction or elimination of expression of the domain or protein. For example, the action of a domain or protein of the present invention can be decreased by blocking or reducing the production of the domain or protein, "knocking out" the gene or portion thereof encoding the domain or protein, reducing domain or protein activity, or inhibiting the activity of the domain or protein. Blocking or reducing the production of a domain or protein can include placing the gene encoding the domain or protein under the control of a promoter that requires the presence of an inducing compound in the growth medium. By establishing conditions such that the inducer becomes depleted from the medium, the expression of the gene encoding the domain or protein (and therefore, of protein synthesis) could be turned off. Blocking or reducing the activity of domain or protein could also include using an excision technology approach similar to that described in U.S. Patent No. 4,743,546, incorporated herein by reference. To use this approach, the gene encoding the protein of interest is cloned between specific genetic sequences that allow specific, controlled excision of the gene from the genome. Excision could be prompted by, for example, a shift in the cultivation temperature of the culture, as in U.S. Patent No. 4,743,546, or by some other physical or nutritional signal.

In one embodiment of the present invention, a genetic modification includes a modification of a nucleic acid sequence encoding protein or domain of an endogenously (naturally) expressed PUFA PKS system, whereby a microorganism that naturally contains such a system is genetically modified by, for example, classical mutagenesis and selection techniques and/or molecular genetic techniques, include genetic engineering techniques. Genetic engineering techniques can include, for example, using a targeting recombinant

vector to delete a portion of an endogenous gene, or to replace a portion of an endogenous gene with a heterologous sequence. Examples of heterologous sequences that could be introduced into a host genome include sequences encoding at least one functional domain from another PKS system, such as a different PUFA PKS system (bacterial or non-bacterial), a type I PKS system (iterative or modular), a type II PKS system, or a type III PKS system. Other heterologous sequences to introduce into the genome of a host includes a sequence encoding a protein or functional domain that is not a domain of a core PKS system, but which will affect the activity of the endogenous PKS system. For example, one could introduce into the host genome a nucleic acid molecule encoding a phosphopantetheinyl transferase (discussed below). Specific modifications that could be made to an endogenous PUFA PKS system are discussed in detail below.

In another aspect of this embodiment of the invention, the genetic modification includes: (1) the introduction into a homologous or heterologous host cell or organism of a recombinant nucleic acid molecule encoding an amino acid sequence having a biological activity of at least one domain of a PUFA PKS system; and/or (2) the introduction into a host cell or organism of a recombinant nucleic acid molecule encoding a protein or functional domain that affects the activity of a PUFA PKS system. The host can include: (1) a host cell or organism that does not express any PKS system for the production of PUFAs, wherein all functional domains of a PUFA PKS system are introduced into the host cell; (2) a host cell that expresses a PKS system for the production of PUFAs (endogenous or recombinant), wherein at least one additional PUFA PKS domain or protein is introduced into the cell or organism. In other words, the present invention intends to encompass any genetically modified cell or organism (e.g., microorganism or plant), wherein the organism comprises at least one PUFA PKS domain or protein described herein, or has been modified to produce a resynthesized and/or chimeric PUFA PKS domain or protein as described herein.

Therefore, using the guidance provided herein, as well as the description of the PUFA PKS systems described herein and known prior to the invention, gene mixing (or mixing of nucleic acid molecules), for example, by the production of chimeric proteins and/or chimeric PUFA PKS systems as described in detail herein, can be used to extend the range of PUFA products, ratios thereof, and production levels thereof, by an organism expressing the PUFA PKS system. For example, the teachings provided herein can be used to improve the amounts of PUFAs produced, to change the ratio of one PUFA to

another, including the ratio of omega-3 to omega-6 PUFAs, and to extend the range of PUFA PKS products to include EPA, DPA (n-3 or n-6), DHA, ARA, GLA, SDA and others, as well as to produce a wide variety of bioactive molecules, including antibiotics, other pharmaceutical compounds, and other desirable products. The method to obtain these improvements includes not only the mixing of genes from various organisms but also various methods of genetically modifying the PUFA PKS genes and nucleic acid molecules disclosed herein. Knowledge of the genetic basis and domain structure of the PUFA PKS systems as described herein provides a basis for designing novel genetically modified organisms. By way of example, various possible manipulations of the PUFA PKS system are discussed in U.S. Patent Application Publication No. 20020194641, U.S. Patent Application Publication No. 20040235127, and U.S. Patent Application Publication No. 20050100995, *supra* with regard to genetic modification and bioactive molecule production. However, this invention provides novel embodiments regarding the manipulation of PUFA production levels by a host organism and the manipulation of the ratio of PUFAs produced by a host organism.

Accordingly, encompassed by the present invention are methods to genetically modify microbial or plant cells by: genetically modifying at least one nucleic acid sequence in the organism that encodes an amino acid sequence having the biological activity of at least one functional domain of a PUFA PKS system according to the present invention, and/or expressing at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding such amino acid sequence. Various embodiments of such sequences, methods to genetically modify an organism, and specific modifications have been described in detail above. Typically, the method is used to produce a particular genetically modified organism that produces a particular bioactive molecule or molecules.

In one embodiment of the present invention, it is contemplated that a mutagenesis program could be combined with a selective screening process to obtain bioactive molecules of interest. This would include methods to search for a range of bioactive compounds. This search would not be restricted to production of those molecules with *cis* double bonds. The mutagenesis methods could include, but are not limited to: chemical mutagenesis, gene shuffling, switching regions of the genes encoding specific enzymatic domains, or mutagenesis restricted to specific regions of those genes, as well as other methods.

For example, high throughput mutagenesis methods could be used to influence or optimize production of the desired bioactive molecule. Once an effective model system has been developed, one could modify these genes in a high throughput manner. Utilization of these technologies can be envisioned on two levels. First, if a sufficiently selective screen for production of a product of interest (e.g., ARA) can be devised, it could be used to attempt to alter the system to produce this product (e.g., in lieu of, or in concert with, other strategies such as those discussed above). Additionally, if the strategies outlined above resulted in a set of genes that did produce the product of interest, the high throughput technologies could then be used to optimize the system. For example, if the introduced domain only functioned at relatively low temperatures, selection methods could be devised to permit removing that limitation.

It is recognized that many genetic alterations, either random or directed, which one may introduce into a native (endogenous, natural) PUFA PKS system, will result in an inactivation of enzymatic functions. A preferred embodiment of the invention includes a system to select for only those modifications that do not block the ability of the PUFA PKS system to produce a product. For example, the FabB- strain of *E. coli* is incapable of synthesizing unsaturated fatty acids and requires supplementation of the medium with fatty acids that can substitute for its normal unsaturated fatty acids in order to grow (see Metz et al., 2001, *supra*). However, this requirement (for supplementation of the medium) can be removed when the strain is transformed with a functional PUFA PKS system (*i.e.* one that produces a PUFA product in the *E. coli* host - see (Metz et al., 2001, *supra*, Figure 2A). The transformed FabB- strain now requires a functional PUFA-PKS system (to produce the unsaturated fatty acids) for growth without supplementation. The key element in this example is that production of a wide range of unsaturated fatty acids will suffice (even unsaturated fatty acid substitutes, such as branched chain fatty acids). Therefore, in another preferred embodiment of the invention, one can create a large number of mutations in one or more of the PUFA PKS genes disclosed herein, and then transform the appropriately modified FabB- strain (e.g. create mutations in an expression construct containing an ER domain and transform a FabB- strain having the other essential domains on a separate plasmid – or integrated into the chromosome) and select only for those transformants that grow without supplementation of the medium (i.e., that still possessed an ability to produce a molecule that could complement the FabB- defect). Additional screens can be developed to look for particular compounds (e.g. use of GC for

fatty acids) being produced in this selective subset of an active PKS system. One could envision a number of similar selective screens for bioactive molecules of interest.

In one embodiment of invention, a genetically modified organism has a modification that changes at least one product produced by the endogenous PKS system, as compared to a wild-type organism. Novel constructs used to produce such modified organisms, as well as the proteins and organisms produced using such constructs, and the methods associated with such modifications, are all encompassed by the invention.

In one preferred embodiment, a genetically modified organism expresses a PUFA PKS system comprising a genetic modification in a β-hydroxy acyl-ACP dehydrase (DH) domain corresponding to the DH2 domain of *Schizochytrium* or *Thraustochytrium*, wherein the modification alters the ratio of long chain fatty acids, and particularly, the ratio of omega-3 to omega-6 long chain fatty acids, produced by the PUFA PKS system, as compared to in the absence of the modification. In one aspect of this embodiment, the modification is selected from the group consisting of a deletion of all or a part of the domain, a substitution of all or part of the domain with a homologous domain or part thereof from a different organism (*e.g.,* a different organism that naturally produces different ratios and/or amounts of PUFAs), and a mutation of the domain.

More specifically, as illustrated herein, the comparison of the *Schizochytrium* and *Thraustochytrium* PUFA PKS architecture (domain organization) with other PUFA PKS system architecture illustrates nature's ability to alter domain order as well as incorporate new domains to create novel end products, or alter the ratios of end products, for example. In addition, the genes can now be manipulated in the laboratory to create new products, as described in the Examples. The inventors have now demonstrated the ability to harness this ability and use it to create novel organisms with novel PUFA profiles and production amounts. Described herein is the manipulation of PUFA PKS systems in either a directed or random manner to influence the end products. For example, in a preferred embodiment, substitution of a DH (FabA-like) domain or biologically active portion thereof of a first PUFA PKS system, and specifically, the DH2 domain described herein, for the homologous DH domain or biologically active portion thereof in a different, second PUFA PKS system is used to alter the ratio of PUFAs produced by the second PUFA PKS system, and particularly, to manipulate the ratio of omega-3 to omega-6 fatty acids produced by the second PUFA PKS system. A similar result can be achieved by substituting an entire protein or any biologically active portion thereof containing such

DH2 domain (*e.g.,* OrfC from *Thraustochytrium* 23B) from a first PUFA PKS system for the homologous protein or portion thereof in a second PUFA PKS system. While the examples described herein utilize the PUFA PKS systems from *Schizochytrium* and *Thraustochytrium*, the similar manipulation of *any* PKS or PKS-like system for the production of PUFAs by modification of the DH2 protein or DH2-like domain is encompassed by the invention. Such modification can be performed alone or in conjunction with other modifications to a PUFA PKS system.

Accordingly, one embodiment of the present invention comprises a chimeric PUFA PKS system and an organism expressing such chimeric PUFA PKS system. In one aspect, the chimeric PUFA PKS system comprises a first PUFA PKS system, wherein the domain or protein of the first PUFA PKS system that corresponds to the DH2 domain or biologically active portion thereof (*e.g.,* from *Schizochytrium* or *Thraustochytrium* described herein) has been modified or substituted with a DH2 domain or protein or biologically active portion thereof from a second, different PUFA PKS system. By "different PUFA PKS system" is meant a PUFA PKS system from a different strain, species, genus or organism, or even a homologue of a natural or wild-type PUFA PKS system. The goal of producing this chimeric protein is to alter the ratio of PUFAs, and particularly the ratio of omega-3 to omega-6 PUFAS, produced by the PUFA PKS system. Therefore, the selection of the different PUFA PKS system should be based on the selection of a second system producing a different, or desired, ratio of PUFAs than the first PUFA PKS system.

In one aspect of the invention, such a chimeric PUFA PKS system comprises a *Schizochytrium* OrfA (SEQ ID NO:2) and OrfB (SEQ ID NO:4) protein as described herein, and a *Thraustochytrium* OrfC (SEQ ID NO:62) protein as described herein. *Schizochytrium*, *E. coli*, and yeast organisms expressing such chimeric PUFA PKS systems are described in the Examples and are encompassed by the present invention, in addition to plants and plant parts expressing such chimeric PUFA PKS systems. In other embodiments, exemplified in the Examples, chimeric PUFA PKS systems are produced comprising all combinations of the *Schizochytrium* and *Thraustochytrium* OrfsA, B and C.

In another aspect of the invention, a chimeric PUFA PKS system comprises a *Schizochytrium* OrfA (SEQ ID NO:2) and OrfB (SEQ ID NO:4) protein as described herein, and a chimeric OrfC protein (encoded by a nucleic acid sequence represented herein by SEQ ID NO:74, encoded by SEQ ID NO:73). The chimeric OrfC polypeptide is

1493 amino acid residues in length. The DH2 region, defined as amino acids 516–1041 of SEQ ID NO:74, consists of the amino acid sequence of the DH2 region of the *Th.*23B OrfC protein, that is, amino acids 491–1016 of SEQ ID NO:62, which includes all of SEQ ID NO:66 and some flanking amino acid sequence from SEQ ID NO:62. With respect to the remainder of the chimeric OrfC amino acid sequence, residues 1–515 and 1042–1493 of SEQ ID NO:74 are identical to *Schizochytrium* OrfC residues 1–515 and 1051–1502 of SEQ ID NO:6, respectively.

In another embodiment of the invention, a genetically modified cell or organism has been modified to express a PUFA PKS system or portion thereof, including a chimeric PUFA PKS system, wherein the nucleic acid sequence(s) encoding the PUFA PKS system or portion thereof is optimized entirely or in part to utilize the preferred codon usage of the host cell or organism. This embodiment is exemplified below and illustrates how production of a bioactive molecule (*e.g.,* a PUFA) can be increased by making such modifications. This embodiment can be utilized together with the other genetic modifications described herein (*e.g.,* the chimeric PUFA PKS and protein embodiments), to improve production of a bioactive molecule in a host organism.

In one aspect of this embodiment, a chimeric PUFA PKS system comprises a *Schizochytrium* OrfA (SEQ ID NO:2) and OrfB (SEQ ID NO:4) protein as described herein, and a *Thraustochytrium* OrfC (SEQ ID NO:62) protein as described herein, wherein the nucleic acid sequence encoding SEQ ID NO:62 is optimized for the host codon usage. An example of such molecule optimized for expression in *Schizochytrium* is described in the Examples, with such nucleic acid sequence encoding *Thraustochytrium* OrfC (synthetic, or codon-optimized, OrfC) represented herein by SEQ ID NO:70. In another embodiment, *Thraustochytrium* OrfA (SEQ ID NO:39) and/or *Thraustochytrium* OrfB (SEQ ID NO:52) can be combined with any one or more of the *Schizochytrium* OrfsA, B, and/or C, and/or with the *Thraustochytrium* OrfC, for expression in *Schizochytrium*. Again, in this example, the nucleic acid molecule encoding the *Thraustochytrium* OrfA and/or *Thraustochytrium* OrfB can be optimized for the host codon usage. Examples of such molecules optimized for expression in *Schizochytrium* are described in the Examples, with the nucleic acid sequence encoding *Thraustochytrium* OrfA (synthetic, or codon-optimized, OrfA) represented herein by SEQ ID NO:71, and with the nucleic acid sequence encoding *Thraustochytrium* OrfB (synthetic, or codon-optimized, OrfB) represented herein by SEQ ID NO:72.

In another aspect of this embodiment, a chimeric PUFA PKS system comprises a *Schizochytrium* OrfA (SEQ ID NO:2) and OrfB (SEQ ID NO:4) protein as described herein, and a chimeric, *and* partially codon-optimized OrfC protein (encoded by a nucleic acid sequence represented herein by SEQ ID NO:75). The protein encoded by SEQ ID NO:75 is also represented by SEQ ID NO:74, which is described above with respect to SEQ ID NO:73. In this case, however, the portion of the nucleic acid sequence encoding SEQ ID NO:66 (DH2 domain), which is derived from *Thraustochytrium*, is optimized for expression in *Schizochytrium* as described in the Examples.

Other codon-optimized nucleic acid sequences for use in *E. coli*, yeast and plants are described above and below in the Examples.

In another embodiment, a genetically modified organism has been modified by transfecting the organism with a recombinant nucleic acid molecule encoding a protein that regulates the chain length of fatty acids produced by the PUFA PKS system. For example, the protein that regulates the chain length of fatty acids produced by the PUFA PKS system can be a chain length factor that directs the synthesis of C20 units and/or C22 units.

In another embodiment, a genetically modified organism expresses a PUFA PKS system comprising a modification in an enoyl-ACP reductase (ER) domain, wherein the modification results in the production of a different compound as compared to in the absence of the modification. In one aspect of this embodiment, the modification is selected from the group consisting of a deletion of all or a part of an ER domain, a substitution of an ER domain from a different organism for the ER domain, and a mutation of an ER domain.

In one embodiment of the invention, the genetically modified organism produces a polyunsaturated fatty acid (PUFA) profile that differs from the naturally occurring organism without a genetic modification.

Many other genetic modifications useful for producing bioactive molecules will be apparent to those of skill in the art, given the present disclosure, and various other modifications have been discussed previously herein. The present invention contemplates any genetic modification related to a PUFA PKS system as described herein which results in the production of a desired bioactive molecule.

As described above, in one embodiment of the present invention, a genetically modified organism, such as a genetically modified microorganism or plant, includes an

organism which has an enhanced ability to synthesize desired bioactive molecules (products) or which has a newly introduced ability to synthesize specific products (*e.g.*, to synthesize PUFAs, to synthesize a different profile of PUFAs, or to synthesize a specific antibiotic). According to the present invention, "an enhanced ability to synthesize" a product refers to any enhancement, or up-regulation, in a pathway related to the synthesis of the product such that the microorganism or plant produces an increased amount of the product (including any production of a product where there was none before) as compared to the wild-type microorganism or plant, cultured or grown, under the same conditions. Methods to produce such genetically modified organisms have been described in detail above. In one preferred embodiment, the present invention relates to a genetically modified plant or part of a plant (*e.g.,* wherein the plant has been genetically modified to express a PUFA PKS system, including a chimeric PUFA PKS system, described herein), which includes at least the core PUFA PKS enzyme complex and, in one embodiment, at least one PUFA PKS accessory protein, (*e.g.,* a PPTase), so that the plant produces PUFAs. Preferably, the plant is an oil seed plant, wherein the oil seeds or oil in the oil seeds contain PUFAs produced by the PUFA PKS system. Such oils contain a detectable amount of at least one target or primary PUFA that is the product of the PUFA PKS system.

The present inventors have demonstrated the production of PUFAs in a plant that has been genetically modified to express the genes encoding a PUFA PKS system from *Schizochytrium* and a PUFA PKS accessory enzyme, 4'-phosphopantetheinyl transferase (PPTase) (*e.g.,* see U.S. Patent Application Publication No. 20070089199, *supra*). The oils produced by these plants contain significant quantities of both DHA (docosahexaenoic acid (C22:6, n-3)) and DPA (docosapentaenoic acid (C22:5, n-6), which are the predominant PUFAs (the primary PUFAs) produced by the *Schizochytrium* from which the PUFA PKS genes were derived. Significantly, oils from plants that produce PUFAs using the PUFA PKS pathway have a different fatty acid profile than plants that are genetically engineered to produce the same PUFAs by the "standard" pathway described above. In particular, oils from plants that have been genetically engineered to produce specific PUFAs by the PUFA PKS pathway are substantially free of the various intermediate products and side products that accumulate in oils that are produced as a result of the use of the standard PUFA synthesis pathway. This characteristic is discussed in detail below.

More particularly, efforts to produce long chain PUFAs in plants by the "standard" pathway (described above) have taken the same basic approach, which is dictated by this synthesis pathway. These efforts relied on modification of the plants' endogenous fatty acids by introduction of genes encoding various elongases and desaturases. Plants typically produce 18 carbon fatty acids (e.g., oleic acid, linoleic acid, linolenic acid) via the Type II fatty acid synthase (FAS) in its plastids. Often, a single double bond is formed while that fatty acid is attached to ACP, and then the oleic acid (18:1) is cleaved from the ACP by the action of an acyl-ACP thioesterase. The free fatty acid is exported from the plastid and converted to an acyl-CoA. The 18:1 can be esterified to phosphatidylcholine (PC) and up to two more *cis* double bonds can be added. The newly introduced elongases can utilize substrates in the acyl-CoA pool to add carbons in two-carbon increments. Newly introduced desaturases can utilize either fatty acids esterified to PC, or those in the acyl-CoA pool, depending on the source of the enzyme. One consequence of this scheme for long chain PUFA production, however, is that intermediates or side products in the pathway accumulate, which often represent the majority of the novel fatty acids in the plant oil, rather than the target long chain PUFA.

For example, using the standard or classical pathway as described above, when the target PUFA product (i.e., the PUFA product that one is targeting for production, trying to produce, attempting to produce, by using the standard pathway) is DHA or EPA, for example (e.g., produced using elongases and desaturases that will produce the DHA or EPA from the products of the FAS system), a variety of intermediate products and side products will be produced in addition to the DHA or EPA, and these intermediate or side products frequently represent the majority of the products produced by the pathway, or are at least present in significant amounts in the lipids of the production organism. Such intermediate and side products include, but are not limited to, fatty acids having fewer carbons and/or fewer double bonds than the target, or primary PUFA, and can include unusual fatty acid side products that may have the same number of carbons as the target or primary PUFA, but which may have double bonds in unusual positions. By way of example, in the production of EPA using the standard pathway (e.g., see U.S. Patent Application Publication 2004/0172682), while the target PUFA of the pathway is EPA (i.e., due to the use of elongases and desaturases that specifically act on the products of the FAS system to produce EPA), the oils produced by the system include a variety of intermediate and side products including: gamma-linolenic acid (GLA; 18:3, n-6);

stearidonic acid (STA or SDA; 18:4, n-3); dihomo-gamma-linolenic acid (DGLA or HGLA; 20:3, n-6), arachidonic acid (ARA, C20:4, n-6); eicosatrienoic acid (ETA; 20:3, n-9) and various other intermediate or side products, such as 20:0; 20:1 (Δ5); 20:1 (Δ11); 20:2 (Δ8,11); 20:2 (Δ11,14); 20:3 (Δ5,11,14); 20:3 (Δ11,14,17); mead acid (20:3; Δ5,8,11); or 20:4 (Δ5,1,14,17). Intermediates of the system can also include long chain PUFAs that are not the target of the genetic modification (e.g., a standard pathway enzyme system for producing DHA can actually produce more EPA as an intermediate product than DHA).

In contrast, the PUFA PKS synthase of the present invention does not utilize the fatty acid products of FAS systems. Instead, it produces the final PUFA product (the primary PUFA product) from the same small precursor molecule that is utilized by FASs and elongases (malonyl-CoA). Therefore, intermediates in the synthesis cycle are not released in any significant amount, and the PUFA product (also referred to herein as the primary PUFA product) is efficiently transferred to phospholipids (PL) and triacylglycerol (TAG) fractions of the lipids. Indeed, a PUFA PKS system may produce two target or primary PUFA products (e.g., the PUFA PKS system from *Schizochytrium* produces both DHA and DPAn-6 as primary products), but DPA is not an intermediate in the pathway to produce DHA. Rather, each is a separate product of the same PUFA PKS system. Therefore, the PUFA PKS genes of the present invention are an excellent means of producing oils containing PUFAs, and particularly, LCPUFAs in a heterologous host, such as a plant, wherein the oils are substantially free (defined below) of the intermediates and side products that contaminate oils produced by the "standard" PUFA pathway.

Therefore, it is an object of the present invention to produce, via the genetic manipulation of plants as described herein, polyunsaturated fatty acids and, by extension, oils obtained from such plants (e.g., obtained from the oil seeds of such plants) comprising these PUFAs. Examples of PUFAs that can be produced by the present invention include, but are not limited to, DHA (docosahexaenoic acid (C22:6, n-3)), ARA (eicosatetraenoic acid or arachidonic acid (C20:4, n-6)), DPA (docosapentaenoic acid (C22:5, n-6 or n-3)), and EPA (eicosapentaenoic acid (C20:5, n-3)). The present invention allows for the production of commercially valuable lipids enriched in one or more desired (target or primary) PUFAs by the present inventors' development of genetically modified plants through the use of the polyketide synthase system of the present invention, as well as components thereof, that produces PUFAs.

According to the present invention, reference to a "primary PUFA", "target PUFA", "intended PUFA", or "desired PUFA" refers to the particular PUFA or PUFAs that are the intended or targeted product of the enzyme pathway that is used to produce the PUFA(s). For example, when using elongases and desaturases to modify products of the FAS system, one can select particular combinations of elongases and desaturases that, when used together, will produce a target or desired PUFA (e.g., DHA or EPA). As discussed above, such target or desired PUFA produced by the standard pathway may not actually be a "primary" PUFA in terms of the amount of PUFA as a percentage of total fatty acids produced by the system, due to the formation of intermediates and side products that can actually represent the majority of products produced by the system. However, one may use the term "primary PUFA" even in that instance to refer to the target or intended PUFA product produced by the elongases or desaturases used in the system.

When using a PUFA PKS system as preferred in the present invention, a given PUFA PKS system derived from a particular organism will produce particular PUFA(s), such that selection of a PUFA PKS system from a particular organism will result in the production of specified target or primary PUFAs. For example, use of a PUFA PKS system from *Schizochytrium* will result in the production of DHA and DPAn-6 as the target or primary PUFAs. Use of a PUFA PKS system from various Shewanella species, on the other hand, will result in the production of EPA as the target or primary PUFA. It is noted that the ratio of the primary or target PUFAs can differ depending on the selection of the particular PUFA PKS system and on how that system responds to the specific conditions in which it is expressed. For example, use of a PUFA PKS system from Thraustochytrium 23B (ATCC No. 20892) will also result in the production of DHA and DPAn-6 as the target or primary PUFAs; however, in the case of Thraustochytrium 23B, the ratio of DHA to DPAn-6 is about 10:1 (and can range from about 8:1 to about 40:1), whereas in *Schizochytrium*, the ratio is typically about 2.5:1. Therefore, use of a *Thraustochytrium* PUFA PKS system or proteins or domains can alter the ratio of PUFAs produced by an organism as compared to *Schizochytrium* even though the target PUFAs are the same. However, as in detail above, the use of various proteins and domains with proteins and domains from other PUFA PKS systems or other PKS systems (that produce bioactive molecules other than PUFAs) can be combined ("mixed and matched") to produce chimeric proteins and/or chimeric PUFA PKS systems (described above),

resulting in the production of different PUFA profiles, including different PUFA types, amounts, and/or ratios of one PUFA to another.

When using a PUFA PKS system of the present invention, oils produced by the organism, such as a plant, are substantially free of intermediate or side products that are not the target or primary PUFA products and that are not naturally produced by the endogenous FAS system in the wild-type organism (*e.g.,* wild-type plants produce some shorter or medium chain PUFAs, such as 18 carbon PUFAs, via the FAS system, but there will be new, or additional, fatty acids produced in the plant as a result of genetic modification with a PUFA PKS system). In other words, as compared to the profile of total fatty acids from the wild-type plant (not genetically modified) or the parent plant used as a recipient for the indicated genetic modification, the majority of *additional* fatty acids in the profile of total fatty acids produced by plants that have been genetically modified with the PUFA PKS system of the present invention (or a component thereof), comprise the target or intended PUFA products of the PUFA PKS system (*i.e.,* the majority of additional fatty acids in the total fatty acids that are produced by the genetically modified plant are the target PUFA(s)).

According to the present invention, reference to "intermediate products" or "side products" of an enzyme system that produces PUFAs refers to any products, and particularly, fatty acid products, that are produced by the enzyme system as a result of the production of the target or primary PUFA(s) of the system, but which are not the primary or target PUFA(s). In one embodiment, intermediate and side products may include non-target fatty acids that are naturally produced by the wild-type plant, or by the parent plant used as a recipient for the indicated genetic modification, but are now classified as intermediate or side products because they are produced in greater levels as a result of the genetic modification, as compared to the levels produced by the wild-type plant, or by the parent plant used as a recipient for the indicated genetic modification. Intermediate and side products are particularly significant in the standard pathway for PUFA synthesis and are substantially less significant in the PUFA PKS pathway, as discussed above. It is noted that a primary or target PUFA of one enzyme system may be an intermediate of a different enzyme system where the primary or target product is a different PUFA, and this is particularly true of products of the standard pathway of PUFA production, since the PUFA PKS system substantially avoids the production of intermediates. For example, when using the standard pathway to produce EPA, fatty acids such as GLA, DGLA and

SDA are produced as intermediate products in significant quantities (e.g., U.S. Patent Application Publication 2004/0172682 illustrates this point). Similarly, and also illustrated by U.S. Patent Application Publication 2004/0172682, when using the standard pathway to produce DHA, in addition to the fatty acids mentioned above, ETA and EPA (notably the target PUFA in the first example above) are produced in significant quantities and in fact, may be present in significantly greater quantities relative to the total fatty acid product than the target PUFA itself. This latter point is also shown in U.S. Patent Application Publication 2004/0172682, where a plant that was engineered to produce DHA by the standard pathway produces more EPA as a percentage of total fatty acids than the targeted DHA.

Furthermore, to be "substantially free" of intermediate or side products of the system for synthesizing PUFAs, or to not have intermediate or side products present in substantial amounts, means that any intermediate or side product fatty acids (non-target PUFAs) that are produced in the genetically modified plant (and/or parts of plants and/or seed oil fraction) as a result of the introduction or presence of the enzyme system for producing PUFAs (i.e., that are not produced by the wild-type plant or the parent plant used as a recipient for the indicated genetic modification), are present in a quantity that is less than about 10% by weight of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% by weight of the total fatty acids produced by the plant, and more preferably less than about 0.5% by weight of the total fatty acids produced by the plant.

In a preferred embodiment, to be "substantially free" of intermediate or side products of the system for synthesizing PUFAs, or to not have intermediate or side products present in substantial amounts, means that any intermediate or side product fatty acids that are produced in the genetically modified plant (and/or parts of plants and/or in seed oil fraction) as a result of the enzyme system for producing PUFAS (i.e., that are not produced by the wild-type plant or by the parent plant used as a recipient for the indicated genetic modification for production of target PUFAs), are present in a quantity that is less than about 10% by weight of the total additional fatty acids produced by the plant

(additional fatty acids being defined as those fatty acids or levels of fatty acids that are not naturally produced by the wild-type plant or by the parent plant that is used as a recipient for the indicated genetic modification for production of target PUFAs), and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of the total additional fatty acids produced by the plant. Therefore, in contrast to the fatty acid profile of plants that have been genetically modified to produce PUFAs via the standard pathway, the majority of fatty acid products resulting from the genetic modification with a PUFA PKS system will be the target or intended fatty acid products.

When the target product of a PUFA PKS system is a long chain PUFA, such as DHA or DPA (n-6 or n-3) produced by the PUFA PKS system of the invention described herein, intermediate products and side products that are not present in substantial amounts in the total lipids of plants genetically modified with such PUFA PKS can include, but are not limited to: gamma-linolenic acid (GLA; 18:3, n-6); stearidonic acid (STA or SDA; 18:4, n-3); dihomo-gamma-linolenic acid (DGLA or HGLA; 20:3, n-6), arachidonic acid (ARA, C20:4, n-6); eicosatrienoic acid (ETA; 20:3, n-9) and various other intermediate or side products, such as 20:0; 20:1 (Δ5); 20:1 (Δ11); 20:2 (Δ8,11); 20:2 (Δ11,14); 20:3 (Δ5,11,14); 20:3 (Δ11,14,17); mead acid (20:3; Δ5,8,11); or 20:4 (Δ5,1,14,17). In addition, when the target product is a particular PUFA, such as DHA, the intermediate products and side products that are not present in substantial amounts in the total lipids of the genetically modified plants also include other PUFAs, including other PUFAs that are a natural product of a different PUFA PKS system, such as EPA in this example. In some systems, a PUFA PKS system may make more than one PUFA, such as both a C22 and a C20 PUFA, and such combinations of PUFA may represent the target product, while other PUFAs may represent intermediate or side products. It is to be noted that the PUFA PKS system of the present invention can also be used, if desired, to produce as a target PUFA a PUFA that can include GLA, SDA or DGLA (referring to embodiments where oils are produced using components of a PUFA PKS system described herein).

Using the knowledge of the genetic basis and domain structure of the PUFA PKS system described herein, the present inventors have designed and produced constructs

encoding such a PUFA PKS system and have successfully produced transgenic plants expressing the PUFA PKS system. The transgenic plants produce oils containing PUFAs, and the oils are substantially free of intermediate products that accumulate in a standard PUFA pathway (see U.S. Patent Application Publication No. 20070089199, *supra*). The present inventors have also demonstrated the use of the constructs to produce PUFAs in *E. coli*, and also in another eukaryote, yeast, as a proof-of-concept experiment prior to the production of the transgenic plants (U.S. Patent Application Publication No. 20070089199, *supra*). The examples demonstrate that transformation of both yeast and plants with a PUFA PKS system that produces DHA and DPAn-6 as the target PUFAs produces both of these PUFAs as the primary additional fatty acids in the total fatty acids of the plant (*i.e.,* subtracting fatty acids that are produced in the wild-type plant), and in the yeast and further, that any other fatty acids that are not present in the fatty acids of the wild-type plant are virtually undetectable. Specific characteristics of genetically modified plants and parts and oils thereof of the present invention are described in detail elsewhere herein.

Accordingly, one embodiment of the present invention is a method to produce desired bioactive molecules (also referred to as products or compounds) by growing or culturing a genetically modified microorganism or a genetically modified plant of the present invention (described in detail above). Such a method includes the step of culturing in a growth or fermentation medium or growing in a suitable environment, such as soil, a microorganism or plant, respectively, that has a genetic modification as described previously herein and in accordance with the present invention. In a preferred embodiment, the method to produce bioactive molecules of the present invention includes the step of culturing under conditions effective to produce the bioactive molecule a genetically modified organism that expresses a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system as described herein.

In the method of production of desired bioactive compounds of the present invention, a genetically modified microorganism is cultured or grown in a suitable medium, under conditions effective to produce the bioactive compound. An appropriate, or effective, medium refers to any medium in which a genetically modified microorganism of the present invention, when cultured, is capable of producing the desired product. Such a medium is typically an aqueous medium comprising assimilable carbon, nitrogen and

phosphate sources. Such a medium can also include appropriate salts, minerals, metals and other nutrients. Microorganisms of the present invention can be cultured in conventional fermentation bioreactors. The microorganisms can be cultured by any fermentation process which includes, but is not limited to, batch, fed-batch, cell recycle, and continuous fermentation. Preferred growth conditions for potential host microorganisms according to the present invention are well known in the art. The desired bioactive molecules produced by the genetically modified microorganism can be recovered from the fermentation medium using conventional separation and purification techniques. For example, the fermentation medium can be filtered or centrifuged to remove microorganisms, cell debris and other particulate matter, and the product can be recovered from the cell-free supernatant by conventional methods, such as, for example, ion exchange, chromatography, extraction, solvent extraction, membrane separation, electrodialysis, reverse osmosis, distillation, chemical derivatization and crystallization. Alternatively, microorganisms producing the desired compound, or extracts and various fractions thereof, can be used without removal of the microorganism components from the product.

In the method for production of desired bioactive compounds of the present invention, a genetically modified plant or plant part (including a plant cell) is cultured in a growth medium or grown in a suitable medium such as soil, as appropriate. An appropriate, or effective, growth or culture medium has been discussed in detail above. A suitable growth medium for higher plants includes any growth medium for plants, including, but not limited to, soil, sand, any other particulate media that support root growth (e.g. vermiculite, perlite, etc.) or Hydroponic culture, as well as suitable light, water and nutritional supplements which optimize the growth of the higher plant. The genetically modified plants of the present invention are engineered to produce significant quantities of the desired product through the activity of the PUFA PKS system that is genetically modified according to the present invention. The compounds can be recovered through purification processes which extract the compounds from the plant. In a preferred embodiment, the compound is recovered by harvesting the plant. In a particularly preferred embodiment, PUFAs are recovered from the plant or plant part by harvesting the oil from the plant or plant part (e.g., from the oil seeds). In this embodiment, the plant can be consumed in its natural state or further processed into consumable products.

Bioactive molecules, according to the present invention, include any molecules (compounds, products, etc.) that have a biological activity, and that can be produced by a PKS system that comprises at least one amino acid sequence having a biological activity of at least one functional domain of a non-bacterial PUFA PKS system as described herein. Such bioactive molecules can include, but are not limited to: a polyunsaturated fatty acid (PUFA), an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and a cholesterol lowering formulation. One advantage of the non-bacterial PUFA PKS system of the present invention is the ability of such a system to introduce carbon-carbon double bonds in the *cis* configuration, and molecules including a double bond at every third carbon. This ability can be utilized to produce a variety of compounds.

With respect to microorganisms, preferably, bioactive compounds of interest are produced by the genetically modified microorganism in an amount that is greater than about 0.05%, and preferably greater than about 0.1%, and more preferably greater than about 0.25%, and more preferably greater than about 0.5%, and more preferably greater than about 0.75%, and more preferably greater than about 1%, and more preferably greater than about 2.5%, and more preferably greater than about 5%, and more preferably greater than about 10%, and more preferably greater than about 15%, and even more preferably greater than about 20% of the dry weight of the microorganism. For lipid compounds, preferably, such compounds are produced in an amount that is greater than about 5% of the dry weight of the microorganism. Other bioactive compounds, such as antibiotics or compounds that are synthesized in smaller amounts may be produced in quantities known to those of skill in the art, and those strains possessing such compounds are identified as predictably containing a novel PKS system of the type described herein.

In some embodiments, particular bioactive molecules (compounds) are secreted by the microorganism, rather than accumulating in the cells. Therefore, such bioactive molecules are generally recovered from the culture medium and the concentration of the molecule produced will vary depending on the microorganism and the size of the culture., and may be measured in g/L, rather than by dry cell weight.

Preferably, a genetically modified organism (*e.g.,* microorganism or plant) of the invention produces one or more polyunsaturated fatty acids including, but not limited to,

EPA (C20:5, n-3), DHA (C22:6, n-3), DPA (C22:5, n-6 or n-3), ARA (C20:4, n-6), GLA (C18:3, n-6), ALA (C18:3, n-3), and/or SDA (C18:4, n-3)), and more preferably, one or more long chain fatty acids (LCPUFAs), including, but not limited to, EPA (C20:5, n-3), DHA (C22:6, n-3), DPA (C22:5, n-6 or n-3), or DTA (C22:4, n-6). In a particularly preferred embodiment, a genetically modified organism of the invention produces one or more polyunsaturated fatty acids including, but not limited to, EPA (C20:5, n-3), DHA (C22:6, n-3), and/or DPA (C22:5, n-6 or n-3).

Preferably, a genetically modified organism of the invention produces at least one PUFA (the target PUFA), wherein the total fatty acid profile in the organism (or a part of the organism that accumulates PUFAs, such as mature seeds or oil from such seeds, if the organism is an oil seed plant), comprises a detectable amount of this PUFA or PUFAs. Preferably, the PUFA is at least a 20 carbon PUFA and comprises at least 3 double bonds, and more preferably at least 4 double bonds, and even more preferably, at least 5 double bonds. In one embodiment, the PUFA is a PUFA that is not naturally produced by the organism in detectable or significant quantities (*e.g.,* the wild-type organism in the absence of genetic modification, or the parent organism used as a recipient for the indicated genetic modification).

Preferably, the total fatty acid profile in the organism (or part of the organism that accumulates PUFAs) comprises at least 0.1% of the target PUFA(s) by weight of the total fatty acids, and more preferably at least about 0.2%, and more preferably at least about 0.3%, and more preferably at least about 0.4%, and more preferably at least about 0.5%, and more preferably at least about 1%, and more preferably at least about 2 %, and more preferably at least about 3%, and more preferably at least about 4%, and more preferably at least about 5%, and more preferably at least about 10%, and more preferably at least about 15%, and more preferably at least about 20%, and more preferably at least about 25%, and more preferably at least about 30%, and more preferably at least about 35%, and more preferably at least about 40%, and more preferably at least about 45%, and more preferably at least about 50%, and more preferably at least about 55%, and more preferably at least about 60%, and more preferably at least about 65%, and more preferably at least about 70%, and more preferably at least about 75%, and more preferably more that 75% of at least one polyunsaturated fatty acid (the target PUFA) by weight of the total fatty acids, or any percentage from 0.1% to 75%, or greater than 75% (up to 100% or about 100%), in 0.1% increments, of the target PUFA(s). As generally

used herein, reference to a percentage amount of PUFA production is by weight of the total fatty acids produced by the organism, unless otherwise stated (*e.g.,* in some cases, percentage by weight is relative to the total fatty acids produced by an enzyme complex, such as a PUFA PKS system). In one embodiment, total fatty acids produced by a plant are presented as a weight percent as determined by gas chromatography (GC) analysis of a fatty acid methyl ester (FAME) preparation.

As described above, it is an additional characteristic of the total fatty acids produced by the above-described plant (and/or parts of plants or seed oil fraction) that these total fatty acids produced by the plant comprise less than (or do not contain any more than) about 10% by weight of any fatty acids, other than the target PUFA(s) that are produced by the enzyme complex that produces the target PUFA(s). Preferably, any fatty acids that are produced by the enzyme complex that produces the target PUFA(s) (e.g., as a result of genetic modification of the plant with the enzyme or enzyme complex that produces the target PUFA(s)), other than the target PUFA(s), are present at less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% by weight of the total fatty acids produced by the plant.

In another embodiment, any fatty acids that are produced by the enzyme complex that produces the target PUFA(s) other than the target PUFA(s) are present at less than (or do not contain any more than) about 10% by weight of the total fatty acids that are produced by the enzyme complex that produces the target PUFA(s) in the plant (i.e., this measurement is limited to those total fatty acids that are produced by the enzyme complex that produces the target PUFAs), and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% by weight of the total fatty acids, and more preferably less than about 0.5% by weight of the total fatty acids that are produced by the enzyme complex that produces the target PUFA(s) in the plant.

In another aspect of this embodiment of the invention, the total fatty acids produced by the plant (and/or parts of plants or seed oil fraction) contain less than (or do not contain any more than) 10% PUFAs having 18 or more carbons by weight of the total fatty acids produced by the plant, other than the target PUFA(s) or the PUFAs that are present in the wild-type plant (not genetically modified) or in the parent plant used as a recipient for the indicated (initial or sequential) genetic modification. In further aspects, the total fatty acids produced by the plant (and/or parts of plants or seed oil fraction) contain less than 9% PUFAs having 18 or more carbons, or less than 8% PUFAs having 18 or more carbons, or less than 7% PUFAs having 18 or more carbons, or less than 6% PUFAs having 18 or more carbons, or less than 5% PUFAs having 18 or more carbons, or less than 4% PUFAs having 18 or more carbons, or less than 3% PUFAs having 18 or more carbons, or less than 2% PUFAs having 18 or more carbons, or less than 1% PUFAs having 18 or more carbons by weight of the total fatty acids produced by the plant, other than the target PUFA(s) or the PUFAs that are present in the wild-type plant (not genetically modified) or the parent plant used as a recipient for the indicated genetic modification.

In another aspect of this embodiment of the invention, the total fatty acids produced by the plant (and/or parts of plants or seed oil fraction) contain less than (or do not contain any more than) 10% PUFAs having 20 or more carbons by weight of the total fatty acids produced by the plant, other than the target PUFA(s) or the PUFAs that are present in the wild-type plant (not genetically modified) or the parent plant used as a recipient for the indicated (initial or sequential) genetic modification. In further aspects, the total fatty acids produced by the plant (and/or parts of plants or seed oil fraction) contain less than 9% PUFAs having 20 or more carbons, or less than 8% PUFAs having 20 or more carbons, or less than 7% PUFAs having 20 or more carbons, or less than 6% PUFAs having 20 or more carbons, or less than 5% PUFAs having 20 or more carbons, or less than 4% PUFAs having 20 or more carbons, or less than 3% PUFAs having 20 or more carbons, or less than 2% PUFAs having 20 or more carbons, or less than 1% PUFAs having 20 or more carbons by weight of the total fatty acids produced by the plant, other than the target PUFA(s) or the PUFAs that are present in the wild-type plant (not genetically modified) or the parent plant used as a recipient for the indicated genetic modification.

In one embodiment, the total fatty acids in the plant (and/or parts of plants or seed oil fraction) contain less than about 10% by weight of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of a fatty acid selected from any one or more of: gamma-linolenic acid (GLA; 18:3, n-6); stearidonic acid (STA or SDA; 18:4, n-3); dihomo-gamma-linolenic acid (DGLA or HGLA; 20:3, n-6), arachidonic acid (ARA, C20:4, n-6); eicosatrienoic acid (ETA; 20:3, n-9) and various other fatty acids, such as 20:0; 20:1 (Δ5); 20:1 (Δ11); 20:2 (Δ8,11); 20:2 (Δ11,14); 20:3 (Δ5,11,14); 20:3 (Δ11,14,17); mead acid (20:3; Δ5,8,11); or 20:4 (Δ5,1,14,17).

In another embodiment, the fatty acids that are produced by the enzyme system that produces the long chain PUFAs in the plant contain less than about 10% by weight of a fatty acid selected from: gamma-linolenic acid (GLA; 18:3, n-6); stearidonic acid (STA or SDA; 18:4, n-3); dihomo-gamma-linolenic acid (DGLA or HGLA; 20:3, n-6), arachidonic acid (ARA, C20:4, n-6); eicosatrienoic acid (ETA; 20:3, n-9) and various other fatty acids, such as 20:0; 20:1 (Δ5); 20:1 (Δ11); 20:2 (Δ8,11); 20:2 (Δ11,14); 20:3 (Δ5,11,14); 20:3 (Δ11,14,17); mead acid (20:3; Δ5,8,11); or 20:4 (Δ5,1,14,17), as a percentage of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of a fatty acid selected from: gamma-linolenic acid (GLA; 18:3, n-6); stearidonic acid (STA or SDA; 18:4, n-3); dihomo-gamma-linolenic acid (DGLA or HGLA; 20:3, n-6), arachidonic acid (ARA, C20:4, n-6); eicosatrienoic acid (ETA; 20:3, n-9) and various other fatty acids, such as 20:0; 20:1 (Δ5); 20:1 (Δ11); 20:2 (Δ8,11); 20:2 (Δ11,14); 20:3 (Δ5,11,14); 20:3 (Δ11,14,17); mead acid (20:3; Δ5,8,11); or 20:4 (Δ5,1,14,17).

In another embodiment, the fatty acids that are produced by the enzyme system that produces the long chain PUFAs in the plant contain less than about 10% by weight of all of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three

carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds, as a percentage of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of all of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds.

In another embodiment, the fatty acids that are produced by the enzyme system that produces the long chain PUFAs in the plant contain less than about 10% by weight of each of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds, as a percentage of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of each of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds.

In another embodiment, the fatty acids that are produced by the enzyme system that produces the long chain PUFAs in the plant contain less than about 10% by weight of any one or more of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds, as a percentage of the total fatty acids produced by the plant, and more preferably less than about 9%, and more preferably less than about 8%, and more preferably less than about 7%, and more preferably less than about 6%, and more preferably less than about 5%, and more preferably less than about 4%, and more

preferably less than about 3%, and more preferably less than about 2%, and more preferably less than about 1% of any one or more of the following PUFAs: gamma-linolenic acid (GLA; 18:3, n-6), PUFAs having 18 carbons and four carbon-carbon double bonds, PUFAs having 20 carbons and three carbon-carbon double bonds, and PUFAs having 22 carbons and two or three carbon-carbon double bonds.

In one aspect of this embodiment of the invention, the plant produces at least two target PUFAs, and the total fatty acid profile in the plant, or the part of the plant that accumulates PUFAs (including oils from the oil seeds), comprises a detectable amount of these PUFAs. In this embodiment, the PUFAs are preferably each at least a 20 carbon PUFA and comprise at least 3 double bonds, and more preferably at least 4 double bonds, and even more preferably, at least 5 double bonds. Such PUFAs are most preferably chosen from DHA, DPAn-6 and EPA. In one aspect, the plant produces DHA and DPAn-6, and the ratio of DHA to DPAn-6 is from about 1:10 to about 10:1 or greater, including any ratio in between. In a one embodiment, the ratio of DHA to DPA is from about 1:1 to about 3:1, and in another embodiment, about 2.5:1. In one embodiment, the plant produces DHA and EPA.

The invention further includes any seeds produced by the plants described above, as well as any plant parts, oils produced by the plants or seeds produced by the plants. The invention also includes any products produced using the plants, plant parts, seed or oils described herein.

One embodiment of the present invention relates to a method to modify an endproduct containing at least one fatty acid, comprising adding to said endproduct an oil produced by a recombinant host cell that expresses at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system as described herein.

Preferably, the endproduct is selected from the group consisting of a food, a dietary supplement, a pharmaceutical formulation, a humanized animal milk, and an infant formula. Suitable pharmaceutical formulations include, but are not limited to, an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and a cholesterol lowering formulation. In one embodiment, the endproduct is used to treat a condition selected from the group consisting of: chronic inflammation,

acute inflammation, gastrointestinal disorder, cancer, cachexia, cardiac restenosis, neurodegenerative disorder, degenerative disorder of the liver, blood lipid disorder, osteoporosis, osteoarthritis, autoimmune disease, preeclampsia, preterm birth, age related maculopathy, pulmonary disorder, and peroxisomal disorder.

Suitable food products include, but are not limited to, fine bakery wares, bread and rolls, breakfast cereals, processed and unprocessed cheese, condiments (ketchup, mayonnaise, etc.), dairy products (milk, yogurt), puddings and gelatine desserts, carbonated drinks, teas, powdered beverage mixes, processed fish products, fruit-based drinks, chewing gum, hard confectionery, frozen dairy products, processed meat products, nut and nut-based spreads, pasta, processed poultry products, gravies and sauces, potato chips and other chips or crisps, chocolate and other confectionery, soups and soup mixes, soya based products (milks, drinks, creams, whiteners), vegetable oil-based spreads, and vegetable-based drinks.

Yet another embodiment of the present invention relates to a method to produce a humanized animal milk. This method includes the steps of genetically modifying milk producing cells of a milk-producing animal with at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system as described herein.

Methods to genetically modify a host cell and to produce a genetically modified non-human, milk-producing animal, are known in the art. Examples of host animals to modify include cattle, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of a transgene expressing population. For animals, PKS-like transgenes can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUF As in the breast milk of the host animal.

Each publication or reference cited herein is incorporated herein by reference in its entirety.

Any reference to or discussion of any document, act or item of knowledge in this specification is included solely for the purpose of providing a context for the present invention. It is not suggested or represented that any of these matters or any combination thereof formed at the priority date part of the common general knowledge, or was known to be relevant to an attempt to solve any problem with which this specification is concerned.

It is to be noted that, throughout the description and claims of this specification, the word 'comprise' and variations of the word, such as 'comprising' and 'comprises', is not intended to exclude other variants or additional components, integers or steps. Modifications and improvements to the invention will be readily apparent to those skilled in the art. Such modifications and improvements are intended to be within the scope of this invention.

The following examples are provided for the purpose of illustration and are not intended to limit the scope of the present invention.

Examples

Example 1

The following example describes the construction of a synthetic *Th.*23B OrfC cloning vector for use in *Schizochytrium*.

Codon usage data for four large genes from *Schizochytrium* (*e.g.*, ATCC 20888 or *Schizochytrium* N230D) (orfA, orfB, orfC, and FAS; described in U.S. Patent Application Publication No. 20020194641, U.S. Patent Application Publication No. 20070089199, or U.S. Patent Application Publication No. 20050191679) were combined. Given that *Schizochytrium* ATCC 20888 produces high levels of fatty acids, it is expected that these genes are highly expressed. Codons with less than about 3% representation (within those for a given amino acid) were eliminated, and the relative usage of the remaining codons was adjusted. Table 1 shows *Schizochytrium* codon usage, adjusted usage, and codon usage for non-synthetic *Th.*23B orfC. DNA2.0 (Menlo Park, CA) was used to analyze these codon usage data to design and synthesize a coding region for *Thraustochytrium* 23B orfC. Nucleotides were added to both ends of the coding region to encode restriction enzyme recognition sites that would facilitate subsequent manipulation of the synthetic gene. A small number of codons were adjusted (without changing the encoded amino acid of SEQ ID NO:62) to eliminate or add certain restriction enzyme recognition sequences (see below for an example). The resultant synthetic sequence was developed by DNA2.0 within a plasmid vector and is shown in Fig. 2B as "pThOrfC synth". Table 1 shows the codon usage of the synthetic coding region.

Table 1.

amino acid	codon	Schizo A, B & C plus FAS		Adjusted/Target Usage	Th.23B orfC		synthetic Th.23B orfC	
		number	fraction	fraction	number	fraction	number	fraction
Arg	CGG	7	0.013	0	13	0.18	0	0
Arg	CGA	6	0.011	0	13	0.18	0	0
Arg	CGT	94	0.173	0.21	17	0.24	11	0.15
Arg	CGC	436	0.803	0.79	17	0.24	61	0.85
Arg	AGG	0	0.000	0.00	9	0.13	0	0
Arg	AGA	0	0.000	0.00	3	0.04	0	0
Ser	TCG	244	0.327	0.34	19	0.19	32	0.33
Ser	TCA	10	0.013	0.00	16	0.16	0	0
Ser	TCT	64	0.086	0.10	12	0.12	10	0.10
Ser	TCC	230	0.308	0.29	19	0.19	32	0.33
Ser	AGT	19	0.025	0.00	12	0.12	0	0
Ser	AGC	179	0.240	0.27	20	0.20	24	0.24
Leu	CTG	111	0.123	0.13	36	0.28	13	0.10
Leu	CTA	2	0.002	0.00	7	0.05	0	0
Leu	CTT	148	0.164	0.18	33	0.26	33	0.26

amino acid	codon	Schizo A, B & C plus FAS		Adjusted/Target Usage	Th.23B orfC		synthetic Th.23B orfC	
		number	fraction	fraction	number	fraction	number	fraction
Leu	CTC	623	0.690	0.69	27	0.21	82	0.64
Leu	TTG	18	0.020	0.00	21	0.16	0	0
Leu	TTA	1	0.001	0.00	4	0.03	0	0
Gly	GGG	7	0.009	0.00	21	0.18	0	0
Gly	GGA	38	0.047	0.04	33	0.29	5	0.04
Gly	GGT	174	0.216	0.25	17	0.15	35	0.30
Gly	GGC	585	0.728	0.71	44	0.38	75	0.65
Val	GTG	198	0.242	0.29	44	0.38	29	0.25
Val	GTA	4	0.005	0.00	14	0.12	0	0
Val	GTT	103	0.126	0.13	34	0.29	18	0.16
Val	GTC	512	0.627	0.58	24	0.21	69	0.59
Ala	GCG	214	0.159	0.17	21	0.18	20	0.17
Ala	GCA	41	0.031	0.00	36	0.31	0	0
Ala	GCT	236	0.176	0.21	33	0.28	25	0.22
Ala	GCC	853	0.635	0.62	26	0.22	71	0.61
Thr	ACG	156	0.297	0.28	19	0.30	21	0.33
Thr	ACA	13	0.025	0.00	8	0.13	0	0
Thr	ACT	71	0.135	0.22	16	0.25	10	0.16
Thr	ACC	285	0.543	0.50	20	0.32	32	0.51
Pro	CCG	195	0.340	0.32	19	0.24	27	0.35
Pro	CCA	12	0.021	0.00	17	0.22	0	0
Pro	CCT	116	0.202	0.27	29	0.37	19	0.24
Pro	CCC	250	0.436	0.41	13	0.17	32	0.41
Ile	ATA	0	0.000	0.00	2	0.03	0	0
Ile	ATT	136	0.298	0.28	40	0.57	16	0.23
Ile	ATC	320	0.702	0.72	28	0.40	54	0.77
Glu	GAG	683	0.912	0.90	47	0.56	77	0.92
Glu	GAA	66	0.088	0.10	37	0.44	7	0.08
Asp	GAT	143	0.237	0.26	33	0.37	22	0.24
Asp	GAC	460	0.763	0.74	57	0.63	68	0.76
Lys	AAG	551	0.960	0.90	40	0.48	73	0.88
Lys	AAA	23	0.040	0.10	43	0.52	10	0.12
Asn	AAT	22	0.062	0.11	12	0.21	6	0.10
Asn	AAC	331	0.938	0.89	46	0.79	52	0.90
Cys	TGT	7	0.050	0.06	12	0.36	4	0.12
Cys	TGC	134	0.950	0.94	21	0.64	29	0.88
Tyr	TAT	13	0.057	0.39	15	0.34	14	0.32

amino acid	codon	Schizo A, B & C plus FAS		Adjusted/Target Usage	Th.23B orfC		synthetic Th.23B orfC	
		number	fraction	fraction	number	fraction	number	fraction
Tyr	TAC	214	0.943	0.61	29	0.66	30	0.68
Phe	TTT	160	0.451	0.47	44	0.62	28	0.39
Phe	TTC	195	0.549	0.43	27	0.38	43	0.61
Gln	CAG	306	0.924	0.90	26	0.47	50	0.91
Gln	CAA	25	0.076	0.10	29	0.53	5	0.09
His	CAT	29	0.173	0.15	10	0.32	7	0.23
His	CAC	139	0.827	0.85	21	0.68	24	0.77
Met	ATG	291	1.00	1	46	1	46	1
Trp	TGG	104	1.00	1	19	1	19	1

As described above, previous work by the present inventors and colleagues (see Example 8 in U.S. Patent Application Publication No. 20050100995) resulted in the creation of a plasmid in which the (non-synthetic) *Th.*23B orfC coding region was cloned between the *Schizochytrium* orfC upstream and downstream non-coding regions such that a "perfect stitch" with the *Th.*23B coding region was generated. Intermediate plasmids in this process can be used to clone the synthetic *Th.*23B orfC coding region (see Figs. 2A and 2B). In order to most easily utilize one of these intermediate constructs, a 283bp nucleotide sequence was designed by the inventors and synthesized by DNA2.0 to create the "perfect stitch" junctions and to utilize restriction sites within the *Schizochytrium* orfC upstream/downstream regions and designed into the synthetic *Th.*23B orfC gene for subsequent cloning reactions. This short DNA sequence was designated "Th23B synth orfC INT" and was contained within the plasmid "pThOrfC stitch INT".

The 283bp "Th23B synth orfC INT" consists of five segments. The first segment consists of the final 102 bp of the *Schizochytrium* orfC upstream (non-coding) region from a *Spe*I site up to but not including the ATG start codon of *Schizochytrium* orfC (see SEQ ID NO:77). The second segment consists of the initial 9 bp of the synthetic *Th.*23B orfC coding region (SEQ ID NO:61) and contains the start ATG overlapping a designed *San*DI site (GGGTCCC). These segments create the upstream "perfect stitch" junction. The third segment is a 6bp *Bam*HI restriction site (GGATCC) that functions as a spacer. The fourth segment consists of the final 45bp of the *Th.*23B orfC coding region (SEQ ID NO:61) from a designed *Cla*I site to the TAA stop codon. The fifth segment consists of

the initial 121bp of the *Schizochytrium* orfC (non-coding) downstream region (not including the stop codon) to a "reverse" *Bsm*I site. The final six nucleotides of the "Th23B synth orfC INT" fragment in the "forward" orientation are 5′>GCATTC>3′. The reverse complement 5′>GAATGC>3′ is the recognition sequence for *Bsm*I. The fourth and fifth segments create the downstream "perfect stitch" junction.

Construction details of the "perfect stitch" version of the synthetic *Th.*23B orfC coding sequence is given below (see also Figs. 2A and 2B).

Step 1 (Fig. 2A). The "Th23B synth orfC INT" fragment from pThOrfC stitch INT was removed by digestion with *Spe*I and *Bsm*I restriction enzymes, and the fragment was purified by agarose gel electrophoresis (GeneClean Turbo kit, QBioGene). Similarly, the large *Spe*I/*Bsm*I vector fragment from pREZ22 (see U.S. Patent Application Publication No. 20050100995), containing about 2000bp each of the *Schizochytrium* orfC upstream and downstream regions separated by a *Bam*HI recognition site spacer cloned into pBlueScriptII SK(+)) was obtained. These two fragments were ligated and transformed into *E. coli* XL-1 Blue (Stratagene, La Jolla, CA). Clones containing the desired plasmid, "pREZ22 orfC INT", were identified by restriction digests and partial DNA sequencing. This plasmid contains the *Schizochytrium* orfC upstream and downstream regions perfectly stitched to the 5-prime and 3-prime regions, respectively, of the synthetic orfC coding region, but is lacking the bulk of the coding region.

Step 2 (Fig. 2B). The bulk of the synthetic *Th.*23B orfC coding region was obtained from "pThOrfC synth" by digestion with *San*DI and *Cla*I restriction enzymes and purification of the desired DNA fragment (as above). This fragment was ligated into a similarly obtained vector fragment from pREZ22 orfC INT and cloned into *E. coli* (as above). The resulting plasmid, "pThOrfC-synPS", contains the full length synthetic *Th.*23B orfC coding region perfectly stitched to the upstream and downstream regions of the *Schizochytrium* orfC gene. The nucleotide sequence of the coding region of pThOrfC-synPS is represented herein by SEQ ID NO:70. SEQ ID NO:70 encodes SEQ ID NO:62. pThOrfC-synPS has been deposited as ATCC Accession No. PTA-8229, as described previously herein.

Example 2

The following example describes the creation of a construct encoding *Schizochytrium* OrfC comprising a DH2 domain from *Thraustochytrium* 23B.

The DH2 region of *Schizochytrium* ATCC20888 OrfC (SEQ ID NO:30) was replaced with that from *Thraustochytrium* 23B ATCC 20892 (SEQ ID NO:66) at specific 5-prime and 3-prime cross-over points by a combination of PCR-based overlap extension ("Splicing by Overlap Extension" or "SOEing" (Horton, R.M., (1993) *In Vitro* Recombination and Mutagenesis of DNA. SOEing together tailor-made genes. *Methods in molecular Biology* Vol. 15: *PCR Protocols: Current Methods and Applications* Chapter 25 pp 251-266 (B.A. White, Ed.) Humana Press, Totawa, NJ)) and restriction cloning.

More specifically, in this example, the inventors constructed a nucleic acid molecule encoding a hybrid (chimeric) OrfC polypeptide (amino acid sequence represented herein by SEQ ID NO:74), 1493 amino acid residues in length, in which the DH2 region, defined as amino acids 516–1041 of this hybrid, consists of the amino acid sequence of the DH2 region of the *Th.*23B OrfC protein; that is, amino acids 491–1016 of SEQ ID NO:62, which includes all of SEQ ID NO:66 (described as the DH2 domain of *Thraustochytrium* 23B herein). The remainder of the hybrid OrfC amino acid sequence, residues 1–515 and 1042–1493 of SEQ ID NO:74, are identical to *Schizochytrium* OrfC residues 1–515 and 1051–1502 of SEQ ID NO:6, respectively.

The construction of the plasmid encoding this chimeric protein is illustrated in Figs. 3A-3C.

Step 1 Primers prREZ197 (SEQ ID NO:78) and prREZ198 (SEQ ID NO:79) were used to amplify approximately 1.5Kb of the *Schizochytrium* orfC reading frame upstream of the DH2 region using the unmodified *Schizochytrium* orfC gene as a template:

prREZ197 CATATGGCGCTCCGTGTCAA

prREZ198 GCCAGGAAGCTTTGA**CATGGGGTGCCAGGACATCT**

Primer prREZ197 created an *Nde*I site (underlined) at the start ATG codon. Reverse primer prREZ198 (35mer) contained the 5-prime cross-over point generated by 20bp of homology to *Schizochytrium* OrfC sequence (bold type) and 15bp of homology to *Th.*23B OrfC sequence. PCR conditions: 50μL reaction, 1μL *PfuUltra* polymerase (Stratagene) and 1X PfuUltra buffer, 2% DMSO, 0.5μM each dNTP, 0.4μM each prRZ197 and prRZ198, 10ng template (cloned *Schizochytrium* orfC coding region), 1 min. initial denaturation at 94°C, 20 cycles of 1 min. denaturation at 94°C, 1 min. annealing at 52°C, 90 sec. extension at 72°C, and 10 min. final extension. The PCR product was purified following agarose gel electrophoresis using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA).

Step 2 Primers prREZ199 (SEQ ID NO:80) and prREZ200 (SEQ ID NO:81) were used to amplify the *Th.*23B DH2 region (approximately 1.5Kb) using the *Th.*23B orfC gene as a template.

prREZ199 **TCCTGGCACCCCATG**TCAAAGCTTCCTGGCAACCCTA

prREZ200 AGTATACAGAGGTGCTGACA

Primer prREZ199 (37mer) contained the 5-prime cross-over point generated by 22bp of homology to *Th.*23B orfC (DH2) sequence and 15bp of homology to *Schizochytrium* orfC sequence (bold). These latter 15bp also provided overlap with prREZ198 and thus the PCR product of Step 1. Reverse primer prREZ200 incorporated a natural *Bst*Z17I site in *Th.*23B orfC at the 3-prime cross-over point (underline). PCR conditions and fragment purification were as above except primers prREZ199 and prREZ200 were used with 10ng on cloned *Th.*23B orfC coding region as the template.

Step 3. Overlap extension was used to create the full-length fusion between the 5-prime end of the *Schizochytrium* orfC coding region and the *Th.*23B DH2 region. PCR was performed with the product of Step 1 (prREZ197 x prREZ198) and Step 2 (prREZ199 x prREZ200) as templates and the outside primers prREZ197 and prREZ200. PCR conditions: 50μL reaction, 1μL PfuUltra polymerase (Stratagene) and 1X PfuUltra buffer, 2% DMSO, 0.5μM each dNTP, 0.4μM each prRZ197 and prRZ200, 50ng each PCR product from Steps 1 and 2, 1 min. initial denaturation at 94°C, 20 cycles of 1 min. denaturation at 94°C, 1 min. annealing at 52°C, 3.5 min. extension at 72°C, and 10 min. final extension. The PCR product was purified as in Step 1.

Step 4 The product of the PCR reaction in Step 3 was cloned into pCR-BluntII-TOPO (Invitrogen) and transformed into TOP10 *E. coli* (Invitrogen) using the manufacturer's recommended conditions to create pREZ171. The sequence of the insert DNA was confirmed to be as designed.

Step 5 Using restriction sites in the respective vector sequences, the cloned DNA in pREZ171 was transferred to vector pBC KS(+) (Stratagene) as an *Xba*I/*Spe*I fragment to create pREZ175.

Step 6 Plasmid pREZ175 was digested (linearized) with *Bst*Z17I, then partially digested with *Nde*I. A *ca.* 6Kb fragment representing the fused *Schizochytrium* orfC 5-prime region and *Th.*23B DH2 region was cloned into the pREZ172 *Nde*I/*Bst*Z17I vector fragment creating pREZ177. Plasmid pREZ172 contains the entire *Schizochytrium* orfC coding region cloned into the *E. coli* expression vector pColADuet-1 (Novagen) such that

the start ATG codon incorporates an *Nde*I site. It derives from pREZ101 (see Example 5), and had been modified by site-directed mutagenesis (Quik Change kit, Stratagene) to insert an amino acid-neutral *Bst*Z17I site at the 3-prime cross-over site. Specifically, the TAC tyrosine codon at amino acid position 1051 was modified to TAT.

Step 7 Upon analysis of pREZ177 by DNA sequencing, it was discovered that a single base pair at the *Bst*Z17I site had been deleted. Specifically, the expected <GTATAC> was instead <GTAAC>. To correct this error, a *Pci*I restriction fragment containing the correct *Bst*Z17I cross-over point from pDS26 was used to replace the defective *Pci*I fragment in pREZ177. Plasmid pDS26 contains a hybrid orfC coding region that had been previously created for other purposes. The resulting plasmid, pREZ179, therefore contains an entire orfC coding region that is predominantly from *Schizochytrium* but contains a precise replacement of the DH2 region with that from *Th.*23B (the amino acid sequence represented herein by SEQ ID NO:74). Plasmid pREZ179 further represents a unique tool to study function of the hybrid gene in *E. coli* and provides a starting point for the development of expression vectors for other organisms.

The following additional steps (see Fig. 3C) describe the transfer of the hybrid gene from pREZ179 to a vector for gene replacement in *Schizochytrium*.

Step 8 The (unmodified) *Schizochytrium* orfC coding region plus short portions of upstream and downstream flanking sequences was isolated from pBR002 (a clone of the orfC genomic region) as a *Nhe*I/*Bsp*EI fragment. This fragment was then cloned into the vector portion of *Nhe*I/*Bsp*EI-digested pREZ31 (functionally equivalent to pREZ33 described in U.S. Patent Application Publication No. 20050100995, Example 8). The resulting plasmid, pDS48, contains the (unmodified) *Schizochytrium* orfC coding region plus the same upstream and downstream sequences that have been used to drive gene replacement at the orfC locus.

Step 9 A portion of the hybrid orfC reading frame containing the entire exchanged *Th.*23B DH2 region was isolated from pREZ179 as a *Pst*I/*Pfl*MI fragment. This fragment was cloned into the vector portion of *Pst*I/*Pfl*MI-digested pDS48 to yield pDS49. As a result, plasmid pDS49 contains the hybrid orfC within the same context as pREZ33 (full-length *Th.*23B orfC coding region as a "perfect stitch" gene replacement; see U.S. Patent Application Publication No. 20050100995, Example 8). The nucleotide sequence of the coding region of pDS49 is represented herein by SEQ ID NO:73. SEQ ID NO:73 encodes

SEQ ID NO:74. Plasmid pDS49 was deposited as ATCC Accession No. PTA-8230, as described in detail previously herein.

Example 3

The following example describes the construction of a construct encoding *Schizochytrium* OrfC comprising a DH2 domain from *Thraustochytrium* 23B, wherein the DH2 domain has been resynthesized to be optimized for *Schizochytrium* codon usage.

In this example, the inventors constructed a nucleic acid molecule encoding a hybrid OrfC polypeptide (SEQ ID NO:74), 1493 amino acid residues in length, in which the DH2 region, defined as amino acids 516–1041 of this hybrid, consists of the amino acid sequence of the DH2 region of the *Th.*23B OrfC protein; that is, amino acids 491–1016 of SEQ ID NO:62, which includes all of SEQ ID NO:66 (described as the DH2 domain of *Thraustochytrium* 23B herein). The remainder of the hybrid OrfC amino acid sequence, residues 1–515 and 1042–1493 of SEQ ID NO:74, are identical to *Schizochytrium* OrfC residues 1–515 and 1051–1502 of SEQ ID NO:6, respectively. Moreover, in this construct, the DNA sequence encoding amino acids 516–1041 was derived from the "synthetic gene sequence" for OrfC of *Th.*23B that is contained in plasmid pThOrfC synth and pThOrfC_synPS (see Example 1 and SEQ ID NO:70) and which employs codons that are preferred for gene expression in *Schizochytrium*. The construction details are illustrated in Figs. 4A-4C and described below.

The DNA sequences encoding the DH2 region of the T23B OrfC polypeptide were amplified by PCR (Rxn 59/60) from pThOrfC synth using oligonucleotide primers dhd59 (5> G CAC CCC ATG AGC AAG CTC CCC GGC AAC >3; SEQ ID NO:82) and dhd60 (5> GT ATA CAG AGG CGC AGA CAC GTT GTA AG >3; SEQ ID NO:83). The "forward" or sense-strand primer dhd59 overlaps the DNA sequence encoding amino acid residues 491–501 (WHPMSKLPGNP; positions 491-501 of SEQ ID NO:62) of the *Th.*23B OrfC protein. The "reverse" or antisense-strand primer dhd60 overlaps the DNA sequence encoding amino acid residues 1008–1017 (TYNVSAPLYT; positions 1008-1017 of SEQ ID NO:62) of the *Th.*23B OrfC protein. Primer dhd60 contains two mismatches with the pThOrfC synth sequence which are indicated by the boxed residues in the dhd60 sequence above. These changes created a *Bst*Z17 I restriction endonuclease site, indicated by the double-underlined portion of the dhd60 sequence above, in order to facilitate subsequent cloning steps and also introduced two "silent mutations" into the coding sequence of the hybrid protein: CTT(L) to CTG(L) and TAC(Y) to TAT(Y). This

amplification was carried out in a reaction volume of 40 µl of 1X *PfuUltra*™ HF reaction buffer (Stratagene, LaJolla, CA) containing dhd59 and dhd60 at 0.5 µM each, 200 µM dNTPs, 2 units of *PfuUltra*™ high-fidelity DNA polymerase (Stratagene, LaJolla, CA) and 1 ng of pThOrfC synth DNA. Cycling parameters were: 1X [1 min @ 94°C], 28X [(1 min @ 94°C), (0.5 min @ 60°C), (1.5 min @ 72°C)], 1X [8.5 min @ 72°C], and hold @ 4°C. The reaction was performed in a Perkin Elmer GeneAmp® PCR System 2400 thermocycler (Applied Biosystems, Foster City, CA).

The DNA sequence encoding amino acid residues 331–522 of the hybrid OrfC protein encoded by pREZ179 was amplified by PCR (Rxn 57/58) from pREZ179 using oligonucleotide primers dhd57 (5> C TGC AGC CAG ATG CTC AAG ATG TAC ATG >3; SEQ ID NO:84) and dhd58 (5> G GAG CTT GCT CAT GGG GTG CCA GGA CAT CTC >3; SEQ ID NO:85). The "forward" or sense-strand primer dhd57 overlaps the DNA sequence encoding amino acid residues 330–339 (GCSQMLKMYM; positions 330-339 of SEQ ID NO:74) of the hybrid OrfC protein encoded by pREZ179. The "reverse" or antisense-strand primer dhd58 overlaps the DNA sequence encoding amino acid residues 513–523 (EMSWHPMSKLP; positions 513-523 of SEQ ID NO:74) of the hybrid OrfC protein. The 5' end of the forward primer, dhd57, overlaps the *Pst* I site present in the hybrid OrfC coding sequence contained in pREZ179. This amplification was carried out in a reaction volume of 40 µl of 1X *PfuUltra*™ HF reaction buffer (Stratagene, LaJolla, CA) containing dhd57 and dhd58 at 0.5 µM each, 200 µM dNTPS, 2 units of *PfuUltra*™ high-fidelity DNA polymerase (Stratagene, LaJolla, CA) and 1 ng of pREZ179 DNA. Cycling parameters were: 1X [1 min @ 94°C], 28X [(1 min @ 94°C), (0.5 min @ 60°C), (1.5 min @ 72°C)], 1X [8.5 min @ 72°C], and hold @ 4°C. The reaction was performed in a Perkin Elmer GeneAmp System 2400 thermocycler.

Four microliters of each of the 57/58 and 59/60 reactions were run out on 1.2% agarose gel. DNA bands were observed in each case that were consistent with the expected product sizes: 578 bp for the 57/58 product and 1578 bp for the 59/60 product. These bands were excised from the gel and the DNA recovered from the agarose slices using a QIAquick® Gel Extraction Kit (QIAGEN, Inc. Valencia, CA) according to the vendor protocol. The PCR products were recovered in 40 µl of elution buffer.

The 5′ 20 nucleotides of the reverse primer dhd58 (underlined above) comprise the reverse complement of the 5′ 20 nucleotides of dhd59, also underlined above. As a result, there is a 20 bp identical overlap between the 3' end of the Rxn 57/58 product and the 5′

end of the Rxn 59/60 product and this overlap allows subsequent PCR splicing of these two products by the technique of PCR "Splicing by Overlap Extension" or "SOEing" [Horton, R.M., (1993) *In Vitro* Recombination and Mutagenesis of DNA. SOEing together tailor-made genes. *Methods in molecular Biology* Vol. 15: *PCR Protocols: Current Methods and Applications* Chapter 25 pp 251-266 (B.A. White, Ed.) Humana Press, Totawa, NJ]. This spliced fragment then contains useful restriction sites at (*Bst*Z17 I & *Pst* I) or near (*Bsi*W I) its ends.

The PCR splicing reaction (Rxn 57/60) was performed as follows. A 40 µl reaction volume of 1X *PfuUltra*™ HF reaction buffer contained primers dhd57 and dhd60 each at 0.5 µM, 200 µM dNTPS, 2 units of *PfuUltra*™ high-fidelity DNA polymerase (Stratagene, LaJolla, CA) and 0.8 µl of a 50-fold dilution of each of the gel-purified PCR products 57/58 and 59/60. A series of PCR splicing reactions was performed in which the annealing temperature was varied in 1°C increments between 66-70°C. Other cycling parameters were constant: 1X [1 min @ 98°C], 33X [(1 min @ 98°C), (1 min @ 66-70°C), (2.5 min @ 72°C)], 1X [7.5 min @ 72°C], and hold @ 6°C. The reaction was performed in a RoboCycler® Temperature Cycler (Stratagene, LaJolla, CA). Aliquots of these reactions were run out on 1% agarose gel and it was observed that all reactions contained a product consistent in size with the expected product (2136 bp) but other bands were also observed at all annealing temperatures. Therefore, the 3 reactions with annealings at 67, 68 and 69°C were pooled, run out on a 1% agarose gel and the approximately 2.1 kb band of interest was excised and the DNA fragment recovered using a QIAquick® Gel Extraction Kit (QIAGEN, Inc. Valencia, CA) according to the vendor protocol. Eluted DNA was recovered in 30 µl of elution buffer and cloned into the PCR fragment cloning vector pCR®-Blunt II TOPO® (Invitrogen Corp., Carlsbad, CA) using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen Corp., Carlsbad, CA) according to the vendor protocols. Products of the TOPO cloning reaction were used to transform One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen) according to the vendor protocol. Eight of the resulting transformants were grown overnight and plasmid DNAs were prepared and analyzed by restriction endonuclease digestion and agarose gel electrophoresis. Seven of the eight were found to contain the cloned 2.1 kb PCR product 57/60. The cloned PCR 57/60 product of one isolate was sequenced and shown to exactly match the expected sequence. DNA sequencing was performed by the Biotechnology Resource Center of Cornell University (Ithaca, New York) on a fee for service basis using

the Applied Biosystems Automated 3730 DNA Analyzer, with Big Dye Terminator chemistry and AmpliTaq-FS DNA Polymerase (Applied Biosystems, Foster City, CA). The plasmid containing the sequence-verified insert was designated pDD21 and was used in further construction steps described below.

The DNA segment encoding the *Th.*23B DH2 domain optimized for *Schizochytrium* codon usage was excised from pDD21 and cloned into pREZ179 (see Example 2) so that it replaced the native *Th.*23B DH2 domain coding sequence present in that construct. The resulting plasmid, pDD22, was constructed as follows. Purified pDD21 DNA was digested with *Bsi*WI and *Bst*Z17I (New England BioLabs, Beverly MA) according to the vendor protocols. The reaction was subsequently subjected to treatment using the QIAquick® Spin Purification Procedure and QIAquick® PCR Purification Kit (QIAGEN Inc., Valencia, CA) according to the vendor protocol. The purified digestion products were run out on a 1% agarose gel and the 1940 bp *Bsi*WI - *Bst*Z17I fragment was excised and eluted from the agarose using a QIAEX II Gel Extraction Kit (QIAGEN Inc., Valencia, CA) according to the vendor protocol. Purified pREZ179 DNA was also digested *Bsi*WI and *Bst*Z17I and subsequently treated with Antarctic Phosphatase (New England BioLabs, Beverly, MA) according to the vendor protocol. The phosphatased digestion products were also subjected to treatment using the QUIquick® procedure as described above and run out on a 0.7% agarose gel. The ~6.1 Kb *Bsi*WI - *Bst*Z17I vector fragment was excised from the gel and eluted from the agarose using the QIAEX II Gel Extraction Kit described above. These two fragments were ligated in 1X T4 Ligase Reaction Buffer using T4 Ligase, both from New England BioLabs (Beverly, MA). Liagation products were used to transform One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen) according top the vendor protocol. Plasmids DNA from three of the resulting transformants were analyzed by restriction endonuclease digestion and agarose gel electrophoresis and all three were found to have the structure of the expected recombinant. One plasmid was designated pDD22 and was employed in further constructions.

In order to facilitate the introduction of the DNA encoding the hybrid OrfC containing the *Th.*23B DH2 region encoded by *Schizochytrium*-preferred codons into the *Schizochytrium* genome, a *Pst*I-*Pfl*MI DNA segment spanning the sequence encoding the DH2 region was excised from pDD22 and cloned into pDS48 (see Example 2), a vector designed for gene replacement at sequences at the orfC gene locus in *Schizochytrium*. The

resulting plasmid, pDD24, which was used to in subsequent gene replacements, was constructed as follows. The DNA segment encoding T23B DH2 domain and with optimized codon usage was excised from pDD22 and cloned into pDS48 so that it replaced the native *Schizochytrium* DH2 domain coding sequence present in that construct. Purified pDD22 DNA was digested with *Pst*I, *Pfl*MI and *Cla*I (New England BioLabs, Beverly MA) according to the vendor protocols. Digestion with ClaI cleaved a *Pfl*MI-*Pfl*MI fragment that would otherwise migrate close to the position of the *Pst*I – *Pfl*MI ~3.2 Kb fragment of interest. The reaction was subsequently subjected to treatment using the QIAquick® Spin Purification Procedure and QIAquick® PCR Purification Kit (QIAGEN Inc., Valencia, CA) according to the vendor protocol. The purified digestion products were run out on a 0.7% agarose gel and the ~ 3.2 Kb *Pst*I – *Pfl*MI fragment of interest was excised and eluted from the agarose using a QIAEX II Gel Extraction Kit (QIAGEN Inc., Valencia, CA) according to the vendor protocol. Purified pDS48 DNA was similarly digested with *Pfl*MI and *Pst*I, subjected to the QIAquick® treatment as described above and run out on a 0.7% agarose gel. The ~8.0 Kb *Pst*I - *Pfl*MI vector fragment was excised from the gel and eluted from the agarose using the QIAEX II Gel Extraction Kit described above. These two fragments were ligated in 1X T4 Ligase Reaction Buffer using T4 Ligase, both from New England BioLabs (Beverly, MA). Liagation products were used to transform One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen) according top the vendor protocol. Resulting transformants were grown overnight in liquid culture of LB media containing 100 μg / ml of ampicillin at 30°C. Propagation of these transformants at 37°C in liquid cultures was found to result in plasmid instability under some circumstances. Plasmid DNAs from three of the resulting transformants were analyzed by restriction endonuclease digestion and agarose gel electrophoresis and all three were found to have the structure of the expected recombinant. One plasmid was designated pDD24 and subjected to additional restriction endonuclease analysis and was employed in gene replacement experiments in *Schizochytrium* (see Example 4). The nucleotide sequence of the coding region of pDD24 is represented herein by SEQ ID NO:75. SEQ ID NO:75 encodes SEQ ID NO:74. The plasmid pDD24 was deposited as ATCC Accession No. PTA-8226, as described previously herein.

Example 4

The following example describes the expression of various *Th.* 23B orfC constructs described in Examples 1-3 above in *Schizochytrium*, and the analysis of PUFAs produced by such organisms.

Expression of Variant Th.23B orfC genes in Schizochytrium

Schizochytrium strain B32-Z1 (see above and Example 8 in U.S. Patent Application Publication No. 20050100995), which is a *Schizochytrium* with an exact deletion of the *Schizochytrium* orfC coding region, was transformed with plasmid pThOrfC-synPS (full length synthetic *Th.*23B orfC; see Example 1), pDS49 (non-synthetic *Th.*23B DH2 region; see Example 2), and pDD24 (synthetic *Th.*23B DH2 region; see Example 3) by particle bombardment using techniques previously described (see U.S. Patent Application Publication No. 2003/0166207). Prototrophic Zeocin™-sensitive transformants were obtained. Such transformants arose from double cross-over gene replacement events as confirmed by Southern blot and/or PCR for selected strains.

Briefly, particle bombardment utilized the BioRad (Hercules, CA) Biolistic® PDS-1000/He Particle Delivery System. *Schizochytrium* strains for transformation were grown at 29-30°C in M2B medium (plus DHA where appropriate) on a gyratory platform (200rpm) to OD600 = 1 to 2.5 (BioPhotometer, Eppendorf). Cells were collected by centrifugation (3000rpm, 5 min.) and re-suspended in sterile 7.5g/L Na_2SO_4 to OD600 = 30. A 150µL volume of suspended cells were spread in a circular patch (6cm diameter) on a Petri plate containing M2B agar (without DHA). For growth of PUFA auxotrophs, M2B was supplemented with DHA to 0.25mM from a stock of 25mM DHA in 40% (w/v) randomly methlyated ß-cyclodextrin (CTD Inc, High Springs, FL.). When performing bombardments for complementation of DHA auxotrophy, DHA was omitted from the agar medium. Bombardments were carried out in laminar flow hood using 1100psi rupture discs, a 0.25in gap between the disc retaining cap and the macrocarrier cover lid, and the stopping screen support in the middle position. The target shelf is in the L2 (6cm) position. Petri plates containing bombarded DHA auxotrophic *Schizochytrium* strains were incubated at 29-30°C until (prospective prototrophic) colonies develop (3-5 days). Randomly chosen colonies were streaked to M2B agar plates. After growth, several well-isolated colonies were transferred to M2B plates with and without Zeocin® (50µg/mL). Zeocin-sensitive DHA prototrophs (suggestive of a gene replacement event) were selected for further study.

Growth of Schizochytrium for Fatty Acid Analysis

Erlenmeyer flasks (250mL) containing 50mL of M50-20 medium were inoculated with the contents (1mL) of a cryovial of the indicated strain. The flasks were incubated at 29-30°C on a rotating shaker at 200rpm for 72 hours. Similar flasks containing SSFM medium were inoculated with 0.5mL of the M50-20 culture and incubated as above for 5 days. Cells were harvested by centrifugation (4000g, 5 min) after dilution of the broth with an equal volume of 70% isopropanol. The resulting cell pellets were suspended in an original volume of 35% isopropanol water and re-centrifuged. The washed cell pellets were immediately frozen at -70°C followed by lyophilization. The fatty acid content of the dried biomass was determined by preparing fatty acid methyl esters (FAMEs) using acidic methanol, extracting them into hexane and analyzing by gas-liquid chromatography.

M50-20 Medium

The components per liter of M50-20 medium are as follows: 12.5g NaCl, 2.5g $MgSO_4 \cdot 7H_2O$, 0.5g KCl, 0.05g $CaCl_2$, 20.0g glucose, 20.0g Na-glutamate, 0.4g KH_2PO_4, 1.0g yeast extract, 0.4g $NaHCO_3$, 5ml PII trace metals (200X PII trace metal solution contains per liter: 6.0g Na_2EDTA, 0.29g $FeCl_3 \cdot 6H_2O$, 6.84g H_3BO_3, 0.86g $MnCl_2 \cdot 4H_2O$, 60mg $ZnCl_2$, 26mg $CoCl_2 \cdot 6H_2O$, 52mg $NiSO_4 \cdot 6H_2O$, 2mg $CuSO_4 \cdot 5H_2O$, and 5mg $NaMoO_4 \cdot 2H_2O$, pH 8.0), 1ml PII vitamin mix (1000X PII vitamin mix contains per liter: 100mg thiamin, 0.5mg biotin, and 0.5mg vitamin B_{12}), pH7.0.

SSFM Medium

The components per liter of SSFM medium are as follows: 13.62g Na_2SO_4, 0.72g K_2SO_4, 0.56g KCl, 2.27g $MgSO_4 \cdot 7H_2O$, 0.19g $CaCl_2$, 0.0565g KH_2PO_4, 0.57g $(NH_4)_2SO_4$, 0.13g Na-glutamate, 100mM MES (4-morpholine ethanesulfonic acid) pH 6.0, 50.0g glucose, 0.16mg vitamin B_{12}, 9.75mg thiamin, 3.33mg calcium pantothenate, 10.3mg $FeSO_4 \cdot 7H_2O$, 3.1mg $MnCl_2 \cdot 4H_2O$, 1.93mg $ZnSO_4 \cdot 7H_2O$, 0.04mg $CoCl_2 \cdot 6H_2O$, 0.04mg $NaMoO_4 \cdot 2H_2O$, 2.07mg $CuSO_4 \cdot 5H_2O$, 2.07mg $NiSO_4 \cdot 6H_2O$, 2.0mg citric acid.

M2B Medium

The components of M2B medium are as follows (per liter): glucose 10g, $(NH_4)_2SO_4$ 0.8g, Na_2SO_4 5.0g, $MgSO_4 \cdot 7H_2O$ 2.0g, KH_2PO_4 0.5g, KCl 0.5g, $CaCl_2 \cdot 2H_2O$ 0.1g, vitamin B_{12} 0.05mg, thiamine·HCl, 0.2mg, calcium pantothenate 0.2mg, $FeSO_4 \cdot 7H_2O$ 3.0mg, $MnCl_2 \cdot 4H_2O$ 1.0mg, $ZnSO_4 \cdot 7H_2O$ 0.8mg, $CoCl_2 \cdot 6H_2O$ 0.02mg, $Na_2MoO_4 \cdot 2H_2O$ 0.01mg, $CuSO_4 \cdot 5H_2O$ 0.6mg, $NiSO_4 \cdot 6H_2O$ 0.8mg, MES buffer 0.1M, pH 6.0 (adjusted with NaOH).

PUFA Analysis of Recombinant Schizochytrium Strains

Table 2 shows the total fatty acid, DHA, and DPAn-6 content (expressed as FAME (fatty acid methyl ester)) of *Schizochytrium* ATCC 20888 and derivative strains in which the native orfC coding region is replaced by all or part of the orfC coding region of *Thraustochytrium* 23B (described in Examples 1-3). Replacement of the entire *Schizochytrium* ATCC 20888 orfC coding region with that from *Th.*23B (strain B34-1) results in a higher DHA/DPAn-6 ratio (closer to that of *Th.*23B) but less total PUFA content. That protein expression is the likely cause of lower total PUFA content is demonstrated by use of the codon-optimized (synthetic) *Th.*23B orfC coding region (*e.g.*, in strain B67-5; transformed with pThOrfC_syn-PS) in which PUFA production is increased over wild-type levels while the enhanced DHA/DPAn-6 ratio is maintained. Substitutions of just the *Schizochytrium* DH2 region with that of *Thraustochytrium* show a similar pattern. The strain with the codon-optimized *Th.*23B DH2 region (B69-2; transformed with pDD24) yields higher PUFA than the strain with the non-optimized DH2 region (B105-1A1; transformed with pDS49). However, the DHA/DPA ratio in strain B105-1A1 (non-optimized DH2 region) was notably high.

Interestingly, strain B69-6 produces high levels of DHA and a relatively high DHA/DPA ratio. This strain resulted from the same transformation of strain B32-Z1 with plasmid pDD24 that produced strain B69-2. However, strain B69-6 does not have a correct integration/gene replacement of the modified orfC coding region (as determined by PCR analysis), although the exact nature of the discrepancy is not known.

Given these data, production-scale fermentations can be developed with strain B69-2 to achieve maximal DHA production, or strains B69-6 or B105-1A1 if the greatest DHA/DPA ratio is desired.

Table 2. Summary of orfC Variants

Strain	FAME (%dcw)	DHA (%dcw)	DPAn-6 (%dcw)	DHA (%FAME)	DHA/DPA	strain description
ATCC20888	71.4	16.5	3.64	22.9	4.5	wild type *Schizochytrium*
B34-1	78.4	13.4	1.24	17.0	10.8	(non-synth.) *Th.*23B orfC
B67-5	73.0	21.3	1.85	28.9	11.5	synth. *Th.*23B orfC
B105-1A1	73.5	19.4	1.31	26.4	14.8	(non-synth.) *Th.*23B DH2
B69-2	73.0	23.0	2.31	31.6	10.0	synth. *Th.*23B DH2
B69-6	73.8	22.4	1.76	30.3	12.7	synth. *Th.*23B DH2

Dcw dry cell weight
FAME fatty acid methyl ester

*Th.*23B *Thraustochytrium* sp. 23B; ATCC20892

Example 5

The following example describes the production of DHA and DPA in *E. coli* by a multi-plasmid system, and further illustrates that the DH2 domain of the PUFA PKS system controls the ratio of fatty acid production by the system.

The inventors have previously demonstrated production of DHA and DPA in *E. coli* by the use of T7 inducible system to express OrfA, OrfB*, OrfC from *Schizochytrium* and HetI from *Nostoc* (Example 3, page 41, U.S. Patent Application Publication No. 20050100995). In this previous example, OrfA, OrfB* and OrfC were contained on a single plasmid. In order to create a system more amenable to genetic manipulation, the individual coding regions from *Schizochytrium* were cloned on a set of compatible expression plasmids designed for the coexpression of multiple target genes. The expression of the target genes is similarly driven by the inducible T7 promoter on this Duet series of plasmids (Novagen). *Schizochytrium* orfA was cloned as an *Nde*I – *Xba*I fragment from pBR115L1 into the expression vector pETDuet-1 to create pREZ91 (pBR115L1 is referenced in the generation of the final expression plasmid in Example 3, page 41, U.S. Patent Application Publication No. 20050100995). *Schizochytrium* orfB* was cloned as an *Nde*I – *Xba*I fragment from pJK780 into the expression vector pCDFDuet-1 to create pREZ96 (pJK780 is referenced in the generation of the final expression plasmid in Example 3, page 41, U.S. Patent Application Publication No. 20050100995). *Schizochytrium* orfC was cloned as an *Nde*I- *Xba*I fragment from pJK510 into pColADuet-1 to create pREZ101 (pJK510 is referenced in the generation of the final expression plasmid in Example 3, page 41, U.S. Patent Application Publication No. 20050100995). The required accessory gene *het*I, encoding a phosphopantetheine transferase (PPTase), was supplied on a pACYC184-based plasmid pJK737 previously described (Example 3, page 41, U.S. Patent Application Publication No. 20050100995). OrfA, OrfB*, OrfC and *het*I, contained separately on plasmids pREZ91, pREZ96, pREZ101 and pJK737 respectively, were transformed into *E. coli* stain BLR (DE3) (Novagen) which contains an inducible T7 RNA polymerase.

Production of DHA and DPA was detected in *E. coli* cells grown in Luria Broth (LB) at both 25°C and 30°C (see Table 3 below) using these multi-plasmid strains. Single colonies were inoculated into LB broth supplemented with antibiotics to maintain each

plasmid in the given strain and grown overnight at the desired temperature (25°C or 30°C). Volumes of 300μL of these cultures were then used to inoculate main cultures of 30mL LB with appropriate antibiotics. The main cultures were grown at the indicated temperature until OD600 (BioPhotometer, Eppendorf) was between 0.45 and 0.55, at which point the cultures were induced with IPTG to a final concentration of 1 mM. The cultures were then maintained under these expression conditions for 24 hours after which the cells were collected by centrifugation and prepared for FAME analysis. The typical level of PUFA produced (as percentages of total FAME) at 30°C was 10% DHA and 6% DPA (16% total PUFA) for the strain carrying *Schizochytrium* orfC. The DHA/DPA ratio of 1.7 approximates that seen in *Schizochytrium* (see Table 2 below).

The expression of the *Schizochytrium* genes required for DHA and DPA production in *E. coli* on separate plasmids provided the inventors with the ability to more easily study and manipulate PUFA biosynthetic genes. As described in U.S. Patent Application Publication No. 2005/0100995, Example 8, it was demonstrated that in *Schizochytrium*, the replacement of orfC with the homologous gene from *Thraustochytrium* 23B altered the PUFAs profile with a shift in the DHA to DPA ratio. The similar experiment was carried out with the *E. coli* multi-plasmid expression system described above, in which the *Schizochytrium* orfC expression plasmid (pREZ101) was replaced with a similar *Thraustochytrium* 23B orfC expression plasmid (pREZ142).

To create pREZ142, the *Th.*23B orfC coding region from pREZ31 was cloned as an *Nco*I/*Sal*I fragment into the Duet vector pColADuet-1. Plasmid pREZ31 is a variant of pREZ33, the "perfect stitch" gene replacement vector (described in Example 1 above and in Example 8 of U.S. Patent Application Publication No. 2005/0100995), in which a *Bam*HI restriction site (underlined below) was engineered just upstream of the start ATG (lower case below). This engineering fortuitously created in pREZ31 an *Nco*I restriction site (italicized below) containing the start ATG which was composed of the last two bases of the *Bam*HI site and the first four bases of the *Th.*23B orfC coding region:

GGAT*CCatgG* (SEQ ID NO:86)

The *Sal*I restriction site used in this cloning is native to the *Schizochytrium* orfC downstream region and is about 250bp downstream of the TAA stop codon. This replacement in the *E. coli* expression system or the *Th.*23B orfC for the *Schizochytrium* orfC resulted in an altered PUFA profile with a shift of the DHA to DPA ratio from 1.5 to

6.8 and the total amount of DHA + DPA was reduced from 10% to 4% when strains were grown and induced at 25°C (see Table below).

Hybrid orfC coding regions were generated in order to determine the region or domain of the gene responsible for control the ratio of DHA to DPA. The hybrid orfC in the expression plasmid pREZ179 contains a central DH2 region derived from *Thraustochytrium* 23B orfC and is flanked upstream and downstream by *Schizochytrium* orfC sequences (see Example 2). When pREZ179 was expressed in the above system in place of pREZ101, a DHA to DPA ratio of 6.5 was seen, while the total PUFA amount was 9% when expressed and induced at 25°C (see Table below). This shift in DHA to DPA ratio in the *E. coli* model expression and maintenance of yield indicated that the central DH2 region of orfC controls the most or all of the ratio of DHA to DPA in PUFA biosynthesis. When this construct was then modified with additional flanking DNA and transformed into *Schizochytrium* to replace the native orfC, a similar shift in DHA to DPA ratio was seen as well as no decrease in production (see Example 4). Similarly when the hybrid orfC was expressed in a yeast system, a shift in DHA to DPA ratio was again seen (see Example 6).

Table 3

orfC form (temperature)	orfC plasmid	DHA + DPA	DHA/DPA
Schizochytrium (30°)	pREZ101	16%	1.7
Schizochytrium (25°)	pREZ101	10%	1.5
*Th.*23B (25°)	pREZ142	4%	6.8
*Th.*23B DH2 (25°)	pREZ179	9%	6.5

Use of multiple expression plasmid system

The above examples, in which the *E. coli* and yeast multi-plasmid expression model systems were used to elucidate the role of orfC and, in particular, the DH2 region, in controlling DHA to DPA ratio in PUFA biosynthesis, demonstrates the utility of these heterologous systems. The results seen in *E. coli* and yeast parallel those seen in *Schizochytrium* in terms of relative effect of the orfC source on DHA/DPA ratio. In a similar manner, the multi-plasmid expression model systems in *E. coli* and yeast are described herein to investigate and engineer other aspects of PUFA biosynthesis including PUFA chain length, degree of fatty acid saturation, and positioning of double bonds. These systems will also allow for the easy expression of genes involved in other types of fatty acid modification such as hydroxylation and glycoslylation. In a similar manner, other PUFA biosynthetic genes from a single organism (as has been done for the

Shewanella japonica cluster described in Example 2, U.S. Patent Application Publication No. 2005/0100995) or from more than one organism can be cloned into this *E. coli* system to facilitate study.

Example 6

The following example describes the method by which *Schizochytrium*'s PUFA synthase subunits A, B and C and *Nostoc het*I were expressed in yeast, and further illustrates that the DH2 domain of the PUFA PKS system controls the ratio of fatty acid production by the system.

Part A

Preliminary expression experiments indicated that *Schizochytrium* OrfC and Het I could be produced as full length proteins in yeast using the native coding regions. In contrast, expression of the native coding regions for *Schizochytrium* OrfsA and B did not result in production of detectable amounts of the expected proteins. The problem seemed to be associated with the translation of the mRNA. (Northern blots showed the presence of mRNAs of the correct size.) Accordingly, synthetic versions of those two coding regions were made with the goal of improving their expression in yeast. The amino acid sequences of the proteins encoded by the synthetic genes are identical to those encoded by the native genes (*i.e.,* SEQ ID NO:2 and SEQ ID NO:4). Initial gene design and complete gene synthesis of orfA and orfB were conducted by Blue Heron Biotechnology, Inc. (Bothell, WA). Codon optimization took into consideration the codon preferences of *S. cerevisiae*. The complete sequences of the synthetic coding regions (designated; sOrfA and sOrfB) are listed as SEQ ID NO:35 (sOrfA) and SEQ ID NO:36 (sOrfB). Each synthetic coding region was appended as follows with DNA to facilitate cloning in the yeast transformation vectors:

upstream sequence (SEQ ID NO:87)
AAGCTTGTGCAGTCAAGTGCGCAAAACCATG

downstream sequence (SEQ ID NO:88)
TAACCCGGG**TCTAGA**.

The start and stop codon positions are underlined and the restriction enzyme recognition sites for *Hin*dIII (upstream) and *Xba*I (downstream) are shown in bold.

The *S. cerevisiae* strain InvSC1 (MATa *his3-Δ1*, *leu2*, *trp1-289*, *ura3-52*) (Invitrogen, Carlsbad, CA) was used for these experiments. The strain was maintained and transformed as per supplier's recommendations. Transformants were grown on

glucose solid medium, raffinose broth and galactose induction medium as per the manufacturer's instructions (Invitrogen). All yeast media components were purchased from Q-BIOgene (Carlsbad, CA).

The *Schizochytrium* PUFA synthase genes and *het*I were cloned into the following transformation vectors: pYES-Leu* (sOrfA; SEQ ID NO:35), pYES3-Tryp (sOrfB; SEQ ID NO:36), pYES2/CT (OrfC; SEQ ID NO:5) and pYES-His* (*het*I; SEQ ID NO:33). Creation of these vectors is described in detail below. Some of the vectors and genes were modified to accommodate specific cloning and expression requirements (described in detail below). Appropriate selection media were used, depending on the particular experiment. The genes were cloned in each case behind GAL1 promoter and expression was induced by re-suspension of washed cells in media containing galactose according to guidelines provide by Invitrogen. Cells were grown at 30°C and harvested (by centrifugation) at the indicated times after being transferred to the induction medium. The cell pellets were freeze dried and FAMEs were prepared using acidic methanol, extracted into hexane, and analyzed by GC.

s*OrfA expression construct*: The sOrfA was cloned into a customized vector, pYES-Leu/CT, constructed as follows. A pYES6/CT vector (Invitrogen) was modified by replacing a region of its DNA containing a blasticidin resistance gene with a segment of DNA containing a *leu2* gene (for selection on media lacking leucine). The blasticidin gene was removed by digesting pYES6/CT with *Bgl*II and *Nhe*I and gel purifying the resulting ~4913 bp vector fragment. The *leu*2 gene was obtained from the yeast vector pRS425 (ATCC 77106, GenBank # U03452). The primers PO-Leu5′ (SEQ ID NO:89) and PO-Leu3′ (SEQ ID NO:90) were used in a PCR reaction with pRS425 as template to generate an ~1812 bp DNA fragment (from bp 664 to 2475 of pRS425) which contains the *leu*2 gene.

PO-Leu5′ GACTGCTAGCTTAAGCAAGGATTTTCTTAAC
PO-Leu3′ GACTGGATCCTCCTGATGCGGTATTTTCTCC

Restriction enzyme recognition sites were incorporated into the primers to facilitate cloning (5′ *Nhe*I and 3′ *Bam*HI underlined). The PCR fragment was digested with *Bam*HI and *Nhe*I and ligated to the 4913 bp vector fragment obtained from the pYES6/CT *Bgl*II/*Nhe*I digest to form pYES6-Leu. This vector was digested with *Hin*dIII and *Xba*I in preparation for insertion of sOrfA. The plasmid from Blue Heron containing the sOrfA

and appropriate flanking DNA was digested with *Hin*dIII and *Xba*I. The 8.8 kb fragment with the complete sOrfA was gel purified and ligated to the prepared pYES6-Leu vector to form pBR882 (pYES6-Leu:sOrfA).

s*Orf B expression construct*: The inventors wished to clone the sOrfB into the pYES3 yeast expression vector which has a tryptophan selection marker. Since the pYES3 vector contains a second *Xba*I restriction site (the second site is in the *trp*1 gene), that restriction enzyme could not conveniently be used for introduction of the sOrf B DNA fragment. The region containing the *Xba*I site downstream of the sOrf B was modified to introduce a unique *Not*I site (also available as a gene insertion cloning site in pYES3) as follows. The plasmid containing the sOrfB fragment from Blue Heron was digested with *Hin*dIII and *Xba*I and the resulting 6.2 kb fragment of interest was gel purified. That fragment was ligated into pYES2/CT (Invitrogen) which had been cut with those same enzymes, yielding the plasmid pBR879. This plasmid was opened by cutting at the unique *Xba*I site. The self complementary oligo linker 5'-CTAGGCGGCCGC-3' (SEQ ID NO:91) was used to create a unique *Not*I site (underlined; it also eliminated the *Xba*I site). This yielded the plasmid pJK894. This construct was digested with *Hin*dIII and *Not*I and the resulting 6.2 kb fragment of interest was gel purified. That fragment was ligated into pYES3/CT (Invitrogen) which had been cut with those same enzymes to form pJK908 (pYES3:sOrfB).

OrfC expression construct: The native orfC had previously been cloned in a bacterial expression vector, and this served as the source for the gene for yeast expression. The bacterial vector was pBluescript II KS (Stratagene), and the coding region plus flanking DNA was cloned into the *Eco*RI (5′) and *Xba*I (3′) sites of the vector. The insert DNA included an *Nde*I restriction site as part of the ATG initiation codon and the TAA stop codon just prior to the *Xba*I site. A bacterial ribosomal binding site sequence was included in the region between the *Eco*RI site and the *Nde*I site containing the initiation codon. Prior to cloning in the yeast vector, the ribosome binding site DNA was removed and replaced with DNA appropriate for expression in the yeast system. The pBluescript plasmid harboring orfC was digested with *Eco*RI and *Nde*I and ligated to the oligonucleotide linkers FL5′ (AATTCAA) and FL3′ (TATTG). The resulting plasmid (designated pKCFL) was digested with *Hind*III (just upstream of the EcoRI site in the pBluescript KS polylinker) and *Xba*I to liberate an ~4526 bp fragment. This fragment was ligated to *Hind*III/*Xba*I-digested pYES2/CT to generate: pYES2/ORFCwt (pYES2:OrfC).

HetI construct: The *hetI* gene from *Nostoc*, encoding a PPTase, was cloned into a customized vector, pYES6-His/CT, which was constructed as follows. A pYES6/CT vector (Invitrogen) was modified by replacing a region of its DNA containing a blasticidin resistance gene with a segment of DNA containing a *his3* gene (for selection on media lacking histidine). The blasticidin gene was removed by digesting pYES6/CT with *Bgl*II and *Nhe*I and gel purifying the resulting ~4913 bp vector fragment. The *his3* gene was amplified from the yeast vector pRS423 (ATCC 77104, GenBank # U03454) using the primers PO-His5′ (SEQ ID NO:92) and PO-His3 (SEQ ID NO:93).

PO-His5′ GACTACTAGTCTAAGAAACCATTATTATCAT
PO-His3′ GACTGGATCCAGCTTTAAATAATCGGTGTCA

This generated an ~1251 bp region of the pRS423 plasmid that contained the *his3* gene. Restriction enzyme recognition sites were incorporated into the primers to facilitate cloning (5′ *Spe*I, and 3′ *Bam*HI, underlined). The PCR fragment was digested with *Spe*I and *Bam*HI and ligated to the ~4913 bp vector fragment obtained from pYES6/CT to form pYES6-His. This vector was digested with *Bam*HI and *Xba*I in preparation for insertion of the *hetI* gene.

The *hetI* gene had previously been cloned and used with the *Schizochytrium* PUFA synthase genes for PUFA production in *E. coli* (U.S. Patent Application Publication No. 20040235127, Example 2). As indicated in that application, there are no methionine codons present in the open reading frame, but there are several potential alternative start codons (TTG and ATT) near the 5′ end (Black and Wolk, 1994, JBC 176, 2282 – 2292). PCR was used to amplify the Orf from *Nostoc* genomic DNA. The 5′ primer was designed so that the first T of the furthest 5′ TTG codon was replaced with an A to create a methionine codon (ATG). The 3′ primer included the TGA stop codon. The amplified region extended from the bp 3994 to 3282 of the *Nostoc* nucleotide sequence deposited as GenBank # L22883 (with nucleotide 3994 being the second T of the TTG codon altered to form the ATG codon). This amplified *hetI* Orf was cloned in a pACYC184 vector along with flanking regulatory elements for expression in *E. coli*. This clone of the *hetI* Orf was used as template DNA to amplify the gene in preparation for cloning into pYES6-His. The primers HetI 5′ (SEQ ID NO:94) and HetI 3′ (SEQ ID NO:95) were used to create a 740bp fragment containing the *hetI* Orf.

HetI 5′ GACTGGATCCGCCACC**ATG**TTGCAGCATACTTGGCTACCAAAACCC

HetI 3′ GACTTCTAGA**TCA**ATAATGCCAGAATTTTGGCTGC

Restriction enzyme recognition sites were incorporated into the primers to facilitate cloning (5′ *Bam*HI and 3′ *Xba*I, underlined). The ATG methionine start codon (5′ primer) and the TGA stop codon (shown as the reverse TCA triplet in the 3′ primer) are shown in bold. The PCR product was digested with *Bam*HI and *Xba*I and ligated into the previously prepared pYES6-His vector to form pYES-His/Het/CT (pYES6-His:HetI).

Results of expressing pYES6-Leu:sOrfA, pYES3:sOrfB, pYES2:OrfC and pYES6-His:HetI in yeast.

Fig. 7 shows a comparison of GC profiles of FAMEs derived from yeast cells expressing the *Schizochytrium* PUFA synthase system (sOrfA, sOrfB, OrfC and *het*I) and one obtained from control cells (lacking the sOrfA gene), such yeast strains denoted herein as strains BRY 4.5 and BRY 3.3, respectively. Cells were collected ~20 hrs after induction. It can be seen that two novel FAME peaks have appeared it the profile of the strain expressing the complete PUFA synthase system. These two peaks were identified as DPAn-6 and DHA by comparison of the elution time with authentic standards and subsequently by MS analyses. As predicted from our characterization of the *Schizochytrium* PUFA synthase, aside from DHA and DPAn-6, no other novel peaks are evident in the profile. Fig. 8 shows the region of the GC chromatogram of Fig. 8 which contains the PUFA FAMEs. Both the control cells and the cell expressing the PUFA synthase contain a peak that elutes near the DHA FAME. This has been identified as C26:0 FAME (by Mass Spectrum analysis) and is likely derived from sphingolipids. Although it elutes close to the DHA peak, the resolution is sufficient so that it does not interfere with the quantitation of DHA. The DPAn-6 peak is well separated from other endogenous yeast lipids in the FAME profile. In this particular example of strain BRY 4.5, the cells expressing the *Schizochytrium* PUFA synthase system accumulated 2.4% DHA and 2.0% DPAn-6 (as a percentage of the total FAMEs; see Table 4 below). The sum of DHA and DPAn-6 is 4.4% of the measured fatty acids in the cells. The ratio of DHA to DPAn-6 observed in the cells was ~1.2:1.

The results presented above showing expression of the *Schizochytrium* PUFA synthase in yeast provide a confirmation of the pathway proposed in the previous applications as well as the predictions in terms of the alterations to the fatty acid profiles that can be expected in yeast and also in plants.

Part B

Expression of Schizochytrium's PUFA synthase Orfs A, B and Nostoc Het I in yeast in combination with a hybrid gene encoding a OrfC containing a DH2 region derived from the orfC homolog of Thraustochytrium 23B, and the effects on the PUFAs produced in those cells.

Expression of hybrid Schizochytrium/Th.23B OrfC genes in yeast: As described in other sections of this application, the inventors have discovered that the main determinants of the ratio of n-3 to n-6 PUFA products of PUFA synthases reside in the OrfC protein and more specifically in the DH2 region of that protein. Gene replacement experiments in both *E. coli* and in *Schizochytrium* using the OrfC homolog derived from Th.23B in combination with the *Schizochytrium*-derived PUFA synthase genes resulted in alteration of the DHA to DPAn-6 ratio produced by those mixed systems. In *E. coli*, the products of the PUFA synthase accumulate as free fatty acids with presumably no influence on the accumulation of the primary products of the enzyme by lipid synthesis enzymes of the host organism. In *Schizochytrium*, the PUFA products accumulate in the esterified lipids, but the endogenous lipid synthesis enzymes are likely to be able to readily accommodate both DHA and DPAn-6 since those are major components of the lipid fraction of the unmodified host. Expression of the mixed PUFA synthase system in yeast would provide a model for heterologous eukaryotic hosts (*e.g.,* plants).

Attempts to express the non-synthetic or fully synthetic *Th.*23B orfC genes in yeast were unsuccessful, as the expected proteins could not be detected. In contrast, expression of the hybrid orfC constructs (described below) resulted in production of active proteins.

Hybrid Schizochytrium / Th.23B OrfCs in pYES2: The plasmid containing the native *Schizochytrium* orfC, pYES2:OrfC (described above), was digested with *Bsi*WI and *Pml*I to remove the section of DNA encoding the DH2 region and some flanking DNA. The region removed was from ~1179 bp (the *Bsi*WI site) to ~3256 bp (the *Pml*I site) of the *Schizochytrium* orfC sequence (SEQ ID NO:5). The resulting 8.4 kb fragment (containing the vector as well as the 5′ and 3′ portions of orfC) was gel purified. A previously described plasmid (see Example 2) containing a hybrid *Schizochytrium*/*Th.*23B orfC (pREZ179 = pColA DUET-*Schizo.* orfC-*Th.*23B DH2 hybrid) was digested with *Bsi*WI and *Pml*I and a 2 kb fragment containing the *Th.*23B DH2 region and flanking *Schizochytrium* DNA was gel purified. The two purified fragments were ligated together to form pYES2: OrfC-23BDH2.

A similar strategy was used to create pYES2: OrfC-s23BDH2. In this case the plasmid used as the source for the synthetic *Th.*23B DH2 region (pDD22; see Example 3) was a hybrid orfC in which the DNA encoding the *Th.*23B DH2 domain was derived from a synthetic coding region whose codons had been modified to more closely match the preferences of *Schizochytrium* (see Example 3).

Results of expressing pYES6-Leu:sOrf A, pYES3:sOrf B, pYES6-His:HetI and pYES2:OrfC-23BDH2 or pYES2: OrfC-s23BDH2 in yeast: Table 4 shows the PUFAs produced in yeast expressing hybrid Orf C constructs in conjunction with the *Schizochytrium* subunits A and B and *Nostoc* HetI. As observed above in part A, the only novel peaks detected in these yeast samples were DHA and DPAn-6. Growth conditions and sample preparation were as described above. Only the relevant PUFA data are shown (as FAMEs given as area %). Samples labeled as BRY 4.21 contain the hybrid orfC with the native *Th.*23B DH2 region, while the sample labeled BRY 4.23 contains the hybrid orfC with the *Th.*23B DH2 region derived from the synthetic gene. Two samples (a and b, from independent isolates) were tested for the BRY 4.21 strain while one isolate of the BRY 4.23 strain was tested. Relative to the cells expressing the *Schizochytrium* orfC, those cells expressing either form of the hybrid orfC have a higher DHA/DPAn-6 ratio (an average of ~2.6 for those with the native *Th.*23B DH2 and a value of ~2.9 for the sample with synthetic *Th.*23B DH2). The expression of the hybrid orfC gene in yeast clearly resulted in an increase in the DHA to DPAn-6 ratio relative to yeast expressing the native *Schizochytrium* orfC gene. The fact that the DHA/DPAn-6 ratio in *Th.*23B cells or in *Schizochytrium* expressing the hybrid orf C is much higher (~8-10) indicates that other factors are contributing to the bias towards accumulation of DHA over DPAn-6. The observation that the ratio did increase in yeast indicates that this construct is a useful model for expressing a PUFA synthase system in heterologous eukaryotic hosts (*e.g.,* yeast or plants).

Table 4

Strain	orfC form	DHA	DPAn-6	DHA + DPA	DHA/DPA
BRY 4.5	*Schizo.*	2.4	2.0	4.4	1.2
BRY 4.21a	*Th.*23B DH2	4.30	1.51	5.81	2.85
BRY 4.21b	*Th.*23B DH2	4.36	1.67	6.03	2.61
BRY 4.23	synth. *Th.*23B DH2	2.71	0.92	3.63	2.95

Example 7

The following example demonstrates the production of PUFAs in fermentation scale experiments using various genetically modified *Schizochytrium* strains described in Example 4.

Experiment 1

Using 2-liter fermentors under typical fermentation conditions, two cultures of a wild-type *Schizochytrium* (ATCC 20888) and two cultures of a transgenic *Schizochytrium* (B67-5, having a codon-optimized (synthetic) *Th*.23B orfC coding region in place of the native *Schizochytrium* orfC coding region; see Example 4) were cultivated to compare the fatty acid profiles. Each strain was fermented in a medium containing carbon, nitrogen, phosphorus, salts, trace metals, and vitamins. Each fermentor was inoculated with a typical seed culture, then cultivated for 80 hours, and fed both a carbon source and a nitrogen source during cultivation. The nitrogen source was fed and consumed only during the growth phase, while the carbon source was fed and consumed throughout the fermentation. After 80 hours, samples from each fermentor were centrifuged, lyophilized and analyzed by gas chromatography for fatty acid content.

Typical fermentation conditions:

Temperature:	28 - 30°C
pH:	5.0 - 7.5
agitation:	100 - 300 cps
airflow:	0.25 - 2.0 vvm
glucose:	5 - 35 g/L (concentration)
inoculum:	7.5% - 15%

The results were as shown in Table 5 below:

Table 5

Strain log hour fermentor	Wild-type 20888 80 BN25	Wild-type 20888 80 BN28	Transgenic B67-5 80 BN26	Transgenic B67-5 80 BN27
% 10:0	0.02	0.01	0.01	0.01
% 12:0	0.20	0.18	0.20	0.20
% 13:0	0.00	0.00	0.07	0.00
% 14:0	9.57	8.89	9.76	9.80
% 16:0	33.68	32.58	34.62	34.51
% 16:1	0.13	0.12	0.18	0.17
% 17:0	0.08	0.09	0.07	0.07

Strain log hour fermentor	Wild-type 20888 80 BN25	Wild-type 20888 80 BN28	Transgenic B67-5 80 BN26	Transgenic B67-5 80 BN27
% 18:0	0.78	0.76	0.77	0.76
% 18:1 n-9	0.00	0.00	0.08	0.08
% 18:1 n-7	0.14	0.12	0.11	0.11
% 18:3 n-6	0.14	0.15	0.08	0.08
% 18:3 n-3	0.03	0.04	0.08	0.08
% 20:0	0.09	0.08	0.08	0.08
% 20:3 n-6	0.32	0.33	0.09	0.09
% 20:4 ARA	0.25	0.30	0.10	0.11
% 20:5 EPA	0.36	0.38	0.60	0.60
% 22:5 n-6	14.98	15.37	6.52	6.52
% 22:5 n-3	0.00	0.00	0.21	0.21
% 22:6 DHA	37.32	38.64	44.47	44.58
DHA / DPA	2.49	2.51	6.82	6.84

As shown in Table 5, strain B67-5 containing the synthetic *Thraustochytrium* 23B orfC coding region in place of the native *Schizochytrium* coding region produced more DHA and had a greater ration of DHA to DPAn-6 than the wild-type *Schizochytrium* strain.

Experiment 2

Using 10-liter fermentors under typical fermentation conditions, one culture of a wild-type *Schizochytrium* (ATCC 20888) and one culture of transgenic *Schizochytrium* (B105-1A1; containing a non-codon-optimized (*Thraustochytrium* native) *Th*.23B DH2 coding region in place of the native *Schizochytrium* DH2 region; see Example 4) were cultivated to compare the fatty acid profiles. Each strain was grown in a medium containing carbon, nitrogen, phosphorus, salts, trace metals, and vitamins. Each fermentor was inoculated with a typical seed culture, then cultivated for 72 hours, and fed both a carbon source and a nitrogen source during cultivation. The nitrogen source was fed and consumed only during the growth phase, while the carbon source was fed and consumed throughout the fermentation. After 72 hours, samples from each fermentor were centrifuged, lyophilized and analyzed by gas chromatography for fatty acid content.

Typical fermentation conditions:

Temperature: 28 - 30°C
pH: 5.0 - 7.5
agitation: 100 - 300 cps
airflow: 0.25 - 2.0 vvm
glucose: 5 - 35 g/L (concentration)
inoculum: 7.5% - 15%

The results are shown in Table 6.

Table 6

Strain Log Hour Vessel	Wild-type 20888 72 BN23	Transgenic B105-1A-1 72 BN24
% 10:0	0.00	0.00
% 12:0	0.26	0.28
% 13:0	0.09	0.10
% 14:0	11.36	12.39
% 16:0	37.10	40.02
% 16:1	0.13	0.15
% 17:0	0.07	0.06
% 18:0	0.83	0.86
% 18:1 n-9	0.00	0.11
% 18:1 n-7	0.08	0.08
% 18:3 n-6	0.13	0.05
% 18:3 n-3	0.00	0.00
% 20:0	0.08	0.10
% 20:3 n-6	0.28	0.00
% 20:4 ARA	0.26	0.00
% 20:5 EPA	0.34	0.35
% 22:5 n-6	13.48	4.40
% 22:5 n-3	0.00	0.00
% 22:6 DHA	34.07	39.56
DHA/DPA	2.53	8.98

Table 6 shows that the strain comprising a *Thraustochytrium* 23B DH2 region in place of the *Schizochytrium* DH2 region has a much higher DHA/DPAn-6 ratio, again illustrating the improved DHA ratio achieved by use of chimeric PUFA PKS systems described herein.

Example 8

This example describes the construction and evaluation of all combinations of synthetic codon-optimized *Th.*23B orfA, orf B, and orfC coding regions expressed in *Schizochytrium*.

Detailed descriptions of methods for the exact replacement of the *Schizochytrium* orfC coding region with the *Th.*23B synthetic codon-optimized orfC coding region have been given above (Examples 1 and 4). Those skilled in the art recognize that these techniques can generally be applied to most genes of interest. Those skilled in the art further recognize that such gene designs and replacements can be achieved by variations on these methods or other methods altogether. For example, multiple genes/coding regions can be deleted simultaneously and replaced simultaneously. In *Schizochytrium*, the orfA and ofB genes are found close together ("linked") in the genome separated by an intergenic region (comprising SEQ ID NO:76). These two coding regions (along with the

intergenic region) can be simultaneously deleted by methods analogous to those described previously for orfC (U.S. Patent Application Publication No. 20050100995). Methods similar to those described in Examples 1 and 4 above can then be used to simultaneously create "perfect stitch" replacements of synthetic codon-optimized *Th.*23B orfA and orfB coding regions (including the entire *Schizochytrium* intergenic region) into the *Schizochytrium* orfA/orfB locus. Strains such as B80-1 and B80-20 (Table 7) were created in this way.

In another example, coding region deletions can be created by a "two-step" method in which a plasmid carrying the marked deletion structure plus a second selectable marker initially recombines in its entirety by a single cross-over event into the target locus. Then, the integrant structure "resolves" by a single cross-over event at a site on the opposite side of the deletion structure such that the second selectable marker is lost and the deletion structure remains in place of the original gene structure (Rothstein R., "Targeting, Disruption, Replacement, and Allele Rescue: Integrative DNA Transformation in Yeast", pp281-301 in Methods in Enzymology, vol. 194 (1991), Elsevier/Academic Press, Amsterdam). The precursor to strain B71-1 (Table 7) was created in this manner.

By the methods outlined here, a set of *Schizochytrium* strains in which all combinations of the synthetic (codon-optimized) *Th.*23B orfA, orfB, and orfC coding regions have replaced the cognate *Schizochytrium* coding regions has been created. The set member containing no *Th.*23B genes is the wild type *Schizochytrium* ATCC20888, and the set member containing only the (full length) synthetic codon-optimized *Th.*23B orfC coding region, B67-5, was described in Example 4 and Table 1 above. This set of eight strains was evaluated for fatty acid production during growth in SSFM medium as described in Example 4 above, and the data are given in Table 7.

Plasmid pDD26 contains the full length synthetic *Th.*23B orfA coding region perfectly stitched to the upstream and downstream regions of the *Schizochytrium* orfA gene. The nucleotide sequence of the coding region of pDD26 is represented herein by SEQ ID NO:71. SEQ ID NO:71 encodes SEQ ID NO:39. pDD26 has been deposited as ATCC Accession No. PTA-8411, as described previously herein.).

Plasmid pDD32 contains the full length synthetic *Th.*23B orfB coding region perfectly stitched to the upstream and downstream regions of the *Schizochytrium* orfB gene. The nucleotide sequence of the coding region of pDD32 is represented herein by

SEQ ID NO:72. SEQ ID NO:72 encodes SEQ ID NO:52. pDD32 has been deposited as ATCC Accession No. PTA-8412, as described previously herein.).

The protein products of all three synthetic codon-optimized *Th.*23B orf coding regions function in *Schizochytrium* and successfully interact with other PUFA synthase components regardless of source. Expression of the *Th.*23B OrfC protein (strain B67-5) causes an increase in the DHA/DPA ratio to a value that approximates that in the native *Th.*23B strain, a result previously demonstrated in Example 4. This phenomenon is seen for all combinations expressing the *Th.*23B OrfC protein (B67-5, B79-11, B79-1, and B80-20). Surprisingly, the combination of synthetic codon-optimized *Th.*23B orfC plus synthetic codon-optimized *Th.*23B orfA coding regions (strain B79-1) leads to the highest level of DHA production, while maintaining the high DHA/DPA ratio. The increased DHA production in this *Schizochytrium* strain appears to be due to both the increased n-3/n-6 ratio caused by *Th.*23B OrfC and increased total PUFA production caused by the interaction of *Th.*23B OrfA with *Th.*23B OrfC.

These data demonstrate that components of the PUFA synthase complex from different organisms can successfully co-function and can confer specific characteristics of the source organism to a new host. Furthermore, manipulation of the source and expression levels of PUFA synthase components can lead to novel profiles, higher productivities, and lower costs of target fatty acids.

Table 7

strain	*Th.* 23B orf gene(s)	FAME (%dcw)	DHA (%dcw)	DPA (%dcw)	DHA (%FAME)	DHA/DPA
ATCC20888	(none)	73.9	16.4	5.4	22.1	3.04
B71-1	A	74.2	17.2	5.15	23.2	3.34
B82-3	B	67.9	15.4	4.93	22.7	3.12
B67-5	C	76.2	22.2	2.88	29.2	7.71
B80-1	AB	77.9	12.8	3.20	16.4	4.00
B79-11	BC	79.1	23.4	2.72	29.6	8.60
B79-1	AC	79.0	31.1	2.90	39.4	10.72
B80-20	ABC	77.4	20.9	2.32	27.0	9.01

Each reference cited herein is incorporated herein by reference in its entirety.

While various embodiments of the present invention have been described in detail, it is apparent that modifications and adaptations of those embodiments will occur to those

skilled in the art. It is to be expressly understood, however, that such modifications and adaptations are within the scope of the present invention, as set forth in the following claims.

The claims defining the invention are as follows:

1. A chimeric PUFA PKS system, wherein an FabA-like β-hydroxyacyl-ACP dehydrase-2 (DH2) domain from a first PUFA PKS system is replaced with a DH2 domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system.

2. The chimeric PUFA PKS system of Claim 1, wherein a protein comprising the DH2 domain from the first PUFA PKS system is replaced with a homologous protein comprising the DH2 domain from the second PUFA PKS system.

3. The chimeric PUFA PKS system of Claim 1 or Claim 2, wherein the DH2 domain from the first or second PUFA PKS system comprises a DH2 domain from *Schizochytrium* or *Thraustochytrium*.

4. The chimeric PUFA PKS system of any one of Claims 1 to 3, wherein the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein the second PUFA PKS system is a *Thraustochytrium* PUFA PKS system.

5. The chimeric PUFA PKS system of any one of Claims 1 to 3, wherein the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein OrfC from the *Schizochytrium* PUFA PKS system is replaced with OrfC from a different thraustochytrid.

6. The chimeric PUFA PKS system of any one of Claims 1 to 3, wherein the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein OrfC from the *Schizochytrium* PUFA PKS system is replaced with OrfC from *Thraustochytrium* 23B.

7. The chimeric PUFA PKS system of Claim 6, wherein the OrfC from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage.

8. The chimeric PUFA PKS system of Claim 7, wherein the nucleic acid sequence comprises SEQ ID NO: 70.

9. The chimeric PUFA PKS system of Claim 6, wherein OrfA from the *Schizochytrium* PUFA PKS system is replaced with OrfA from *Thraustochytrium* 23B.

10. The chimeric PUFA PKS system of Claim 9, wherein the OrfA from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage.

11. The chimeric PUFA PKS system of Claim 10, wherein the nucleic acid sequence comprises SEQ ID NO:71.

12. The chimeric PUFA PKS system of Claim 6, wherein OrfB from the *Schizochytrium* PUFA PKS system is replaced with OrfB from *Thraustochytrium* 23B.

13. The chimeric PUFA PKS system of Claim 12, wherein OrfB from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage.

14. The chimeric PUFA PKS system of Claim 13, wherein the nucleic acid sequence comprises SEQ ID NO: 72.

15. The chimeric PUFA PKS system of any one of Claims 1 to 3, wherein the first PUFA PKS system is a *Schizochytrium* PUFA PKS system, and wherein the DH2 domain of OrfC from the *Schizochytrium* PUFA PKS system is replaced with the DH2 domain from *Thraustochytrium* 23B.

16. The chimeric PUFA PKS system of Claim 15, wherein the nucleic acid sequence comprising the DH2 domain from *Thraustochytrium* 23B comprises SEQ ID NO:73.

17. The chimeric PUFA PKS system of Claim 15, wherein the DH2 from *Thraustochytrium* 23B is encoded by a nucleic acid sequence that is optimized for *Schizochytrium* codon usage.

18. The chimeric PUFA PKS system of Claim 17, wherein the nucleic acid sequence comprising the DH2 domain from *Thraustochytrium* 23B comprises SEQ ID NO:75.

19. The chimeric PUFA PKS system of Claim 1, wherein the chimeric PUFA PKS system comprises:

(a) a protein comprising an amino acid sequence that is at least 95% identical to SEQ ID NO: 74;

(b) a protein comprising an amino acid sequence of SEQ ID NO: 74;

(c) SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:74;

(d) SEQ ID NO:39, SEQ ID NO:4 and SEQ ID NO:62; or

(e) SEQ ID NO:39, SEQ ID NO:4 and SEQ ID NO:74.

20. The chimeric PUFA PKS system of Claim 1, wherein the chimeric PUFA PKS system is encoded by nucleic acid molecules comprising:

(a) SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:70;

(b) SEQ ID NO:1, SEQ IDNO:3 and SEQ ID NO:73;

(c) SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:75; or

(d) SEQ ID NO:71, SEQ ID NO:3 and SEQ ID NO:70.

21. A method of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system, comprising expressing the chimeric PUFA PKS system of any one of Claims 1 to 21 in an organism.

22. The method of Claim 21, wherein the chimeric PUFA PKS system is expressed by a microorganism.

23. The method of Claim 22, wherein the microorganism is a *Schizochytrium* or a yeast.

24. The method of Claim 21, wherein the chimeric PUFA PKS system is expressed by a plant.

25. A genetically modified microorganism or plant or part of the plant, comprising a chimeric PUFA PKS system of any one of Claims 1 to 20.

26. A method of increasing the production of PUFAs and of altering the omega-3 to omega-6 ratio of polyunsaturated fatty acids (PUFAs) produced by a first PUFA PKS system, comprising expressing a chimeric PUFA PKS system in an organism, wherein the Fab A- like β-hydroxyacyl-ACP dehydrase-2 (DH2) domain from a first PUFA PKS system is replaced with a DH2 domain from a different, second PUFA PKS system, to produce a chimeric PUFA PKS system that produces a different ratio of omega-3 to omega-6 PUFAs as compared to the first PUFA PKS system, and wherein the DH2 domain from the second PUFA PKS system is optimized for the codon usage of the organism from which the first PUFA PKS system is derived.

27. An isolated nucleic acid molecule encoding a chimeric OrfC protein that is at least 95% identical to SEQ ID NO: 74.

28. The isolated nucleic acid molecule of Claim 27, wherein the isolated nucleic acid molecule comprises a nucleic acid sequence that is at least 95% identical to SEQ ID NO:73.

29. The isolated nucleic acid molecule of Claim 27 or Claim 28, wherein the nucleic acid sequence is optimized for the codon usage of an organism in which the nucleic acid molecule is to be expressed.

30. The isolated nucleic acid molecule of Claim 27 or Claim 28, wherein the nucleic acid sequence is optimized for the codon usage of an organism from which a portion of the chimeric protein is derived.

31. The isolated nucleic acid molecule of Claim 27, wherein the nucleic acid sequence is at least 95% identical to SEQ ID NO: 75.

32. A recombinant nucleic acid molecule comprising the nucleic acid molecule of any one of Claims 27 to 31.

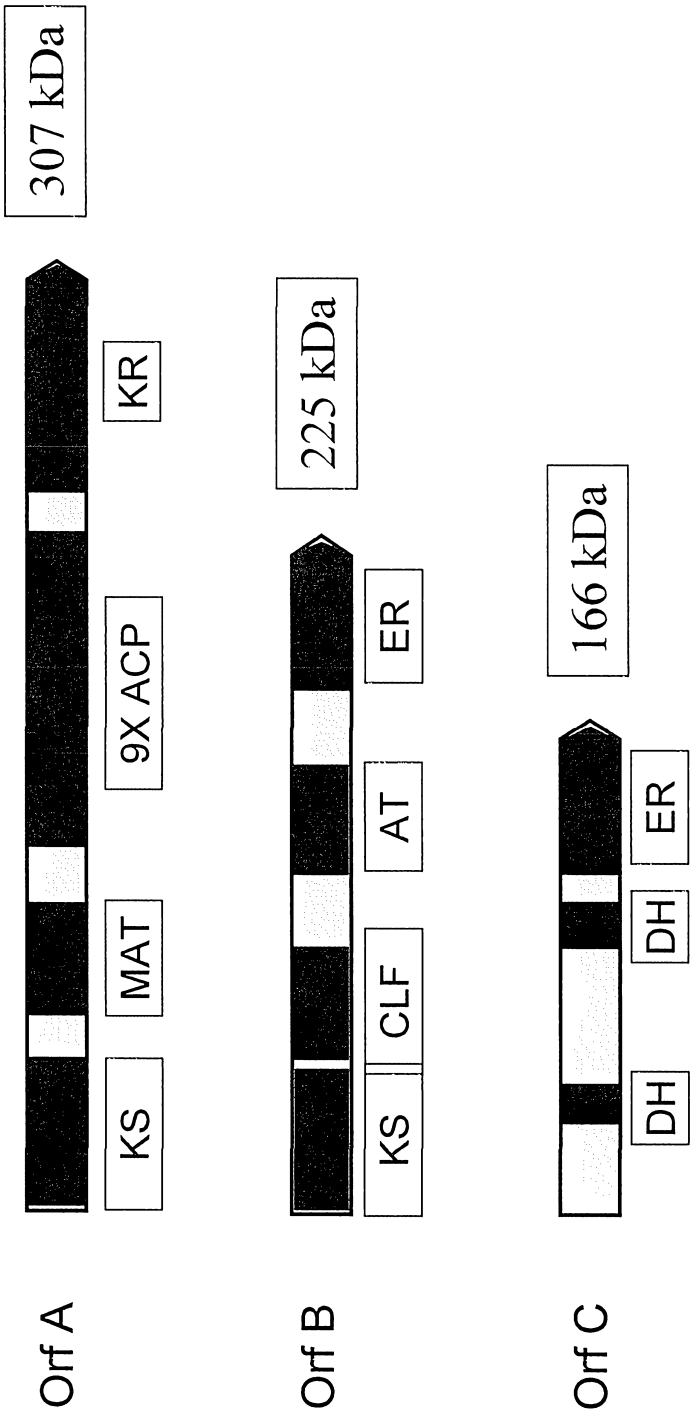
33. A recombinant host cell that has been transfected with the nucleic acid molecule of any one of Claims 27 to 32.

34. The recombinant host cell of Claim 33, wherein the cell is a microorganism or a plant cell.

35. The recombinant host cell of Claim 34, wherein the microorganism is a *Schizochytrium*, a bacterium, or a yeast.

36. A genetically modified plant or part thereof, comprising the plant cell of Claim 34.

FIG. 1


307 kDa
KS
MAT
9X ACP
KR
Orf A
225 kDa
KS
CLF
AT
ER
Orf B
166 kDa
DH
DH
ER
Orf C

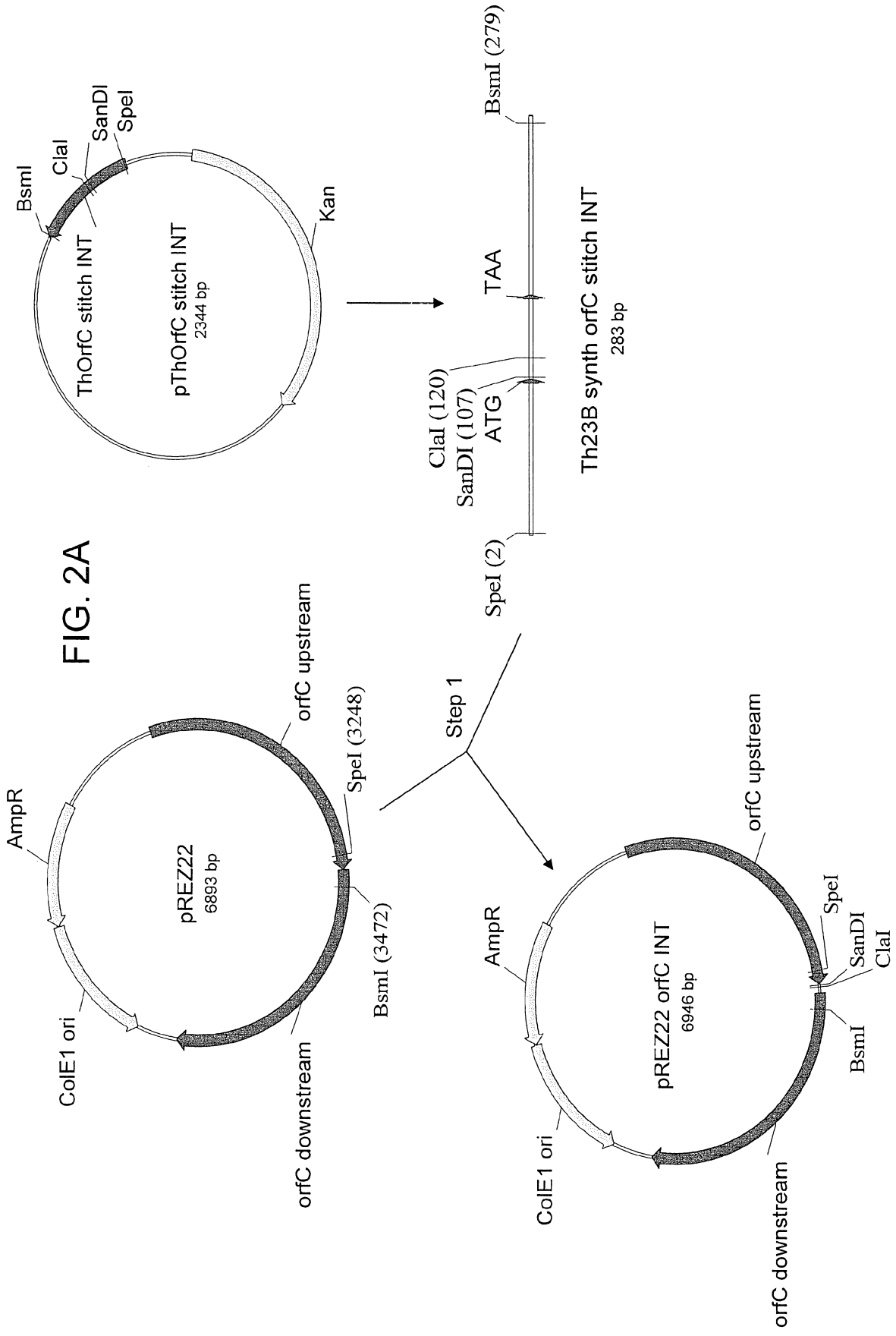

BsmI
ClaI
SanDI
SpeI
ThOrfC stitch INT
pThOrfC stitch INT
2344 bp
Kan
BsmI (279)
TAA
ClaI (120)
SanDI (107)
ATG
SpeI (2)
Th23B synth orfC stitch INT
283 bp
AmpR
pREZ22
6893 bp
orfC upstream
SpeI (3248)
BsmI (3472)
ColE1 ori
orfC downstream
Step 1
AmpR
pREZ22 orfC INT
6946 bp
orfC upstream
SpeI
SanDI
ClaI
BsmI
ColE1 ori
orfC downstream

FIG. 2A

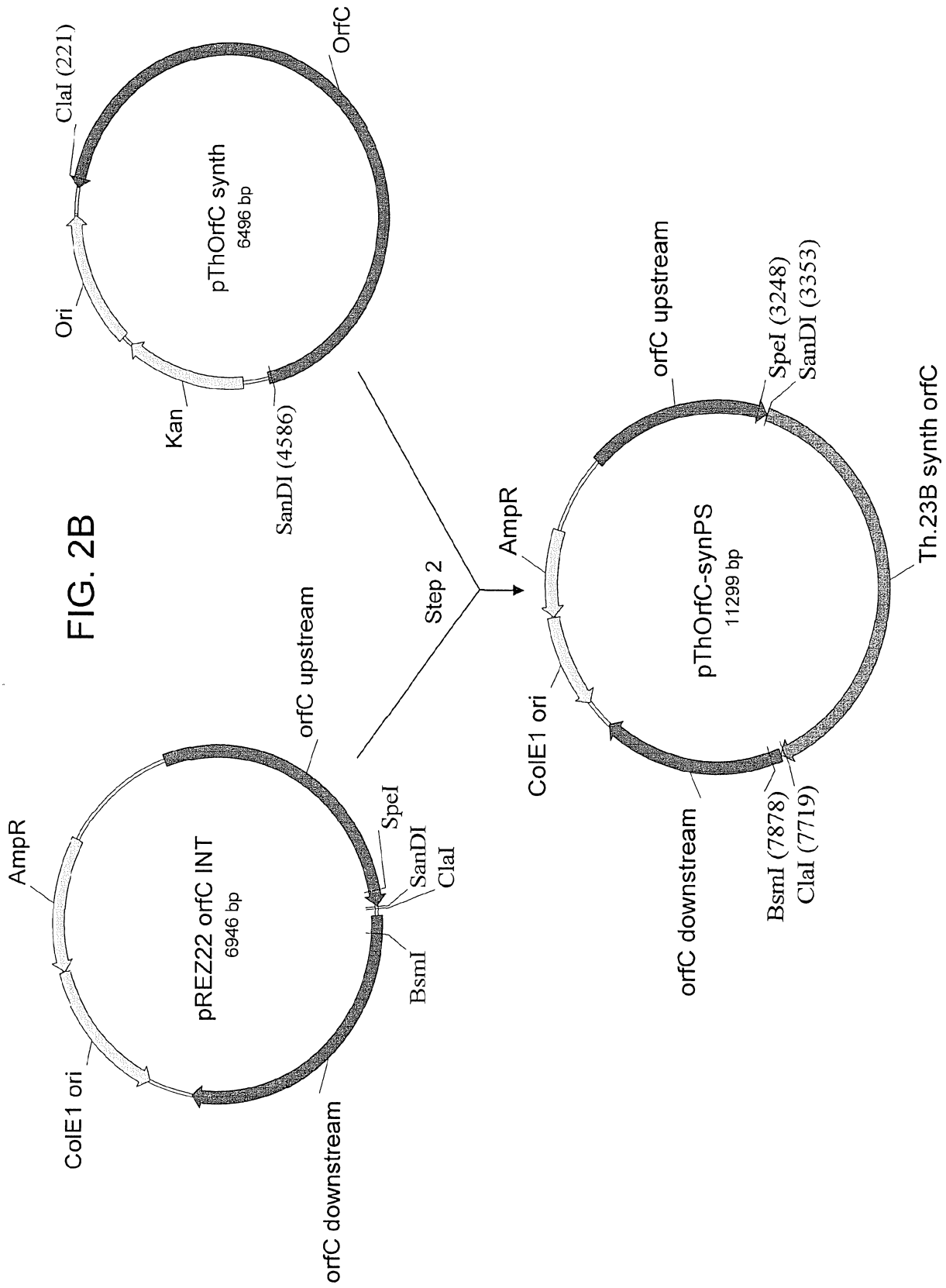
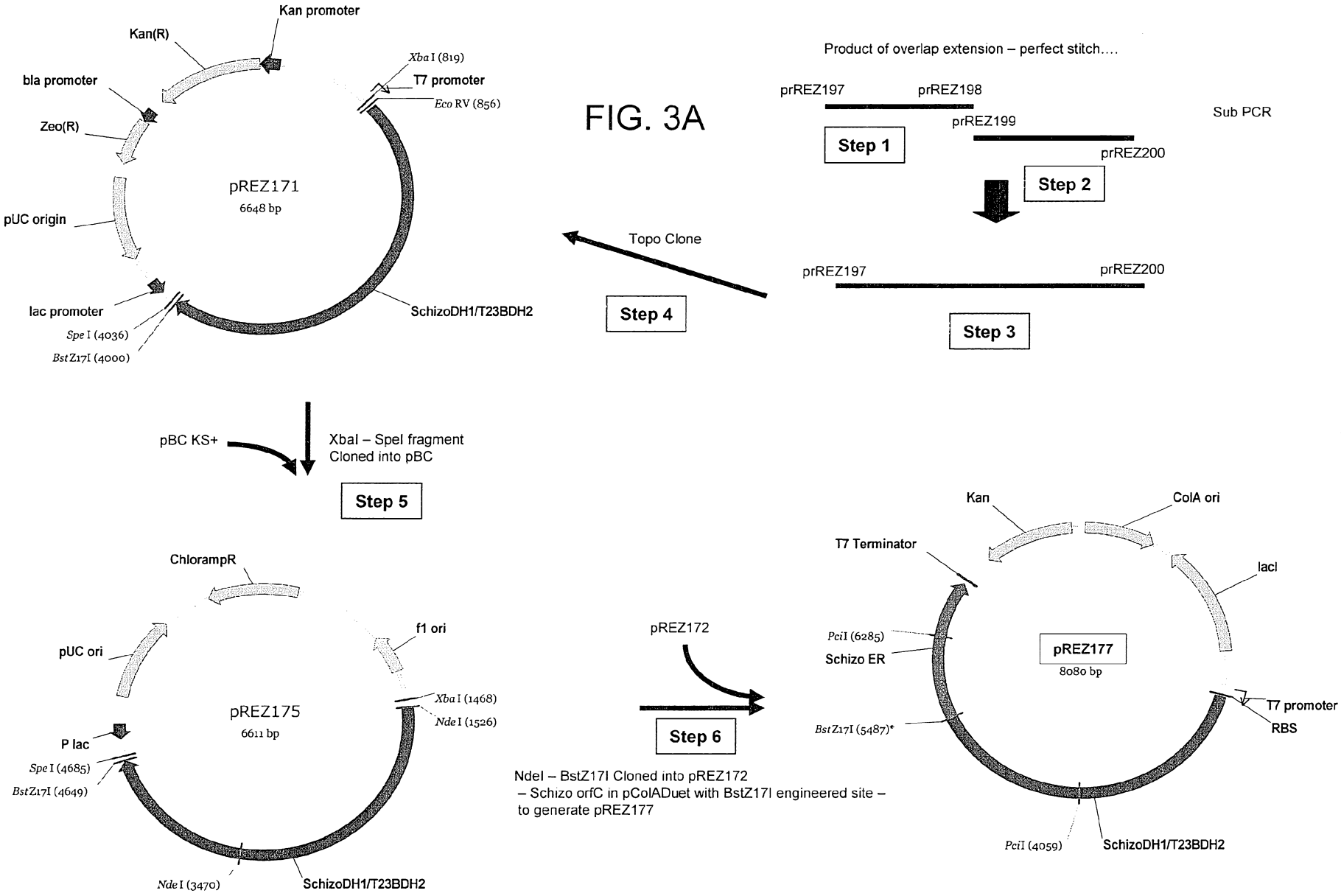

ClaI (221)
OrfC
pThOrfC synth
6496 bp
Ori
Kan
SanDI (4586)
AmpR
pREZ22 orfC INT
6946 bp
ColE1 ori
orfC upstream
SpeI
SanDI
ClaI
BsmI
orfC downstream
Step 2
AmpR
ColE1 ori
pThOrfC-synPS
11299 bp
orfC upstream
SpeI (3248)
SanDI (3353)
Th.23B synth orfC
orfC downstream
BsmI (7878)
ClaI (7719)

FIG. 2B

FIG. 3A


Kan promoter
Kan(R)
bla promoter
Zeo(R)
pUC origin
lac promoter
Spe I (4036)
BstZ17I (4000)
Xba I (819)
T7 promoter
Eco RV (856)
pREZ171
6648 bp
SchizoDH1/T23BDH2
Product of overlap extension – perfect stitch....
prREZ197
prREZ198
prREZ199
prREZ200
Sub PCR
Step 1
Step 2
Step 3
Step 4
Topo Clone
prREZ197
prREZ200
pBC KS+
XbaI – SpeI fragment
Cloned into pBC
Step 5
ChlorampR
pUC ori
P lac
Spe I (4685)
BstZ17I (4649)
f1 ori
Xba I (1468)
Nde I (1526)
pREZ175
6611 bp
Nde I (3470)
SchizoDH1/T23BDH2
pREZ172
Step 6
NdeI – BstZ17I Cloned into pREZ172
– Schizo orfC in pColADuet with BstZ17I engineered site –
to generate pREZ177
Kan
ColA ori
T7 Terminator
lacI
PciI (6285)
Schizo ER
BstZ17I (5487)*
pREZ177
8080 bp
T7 promoter
RBS
PciI (4059)
SchizoDH1/T23BDH2

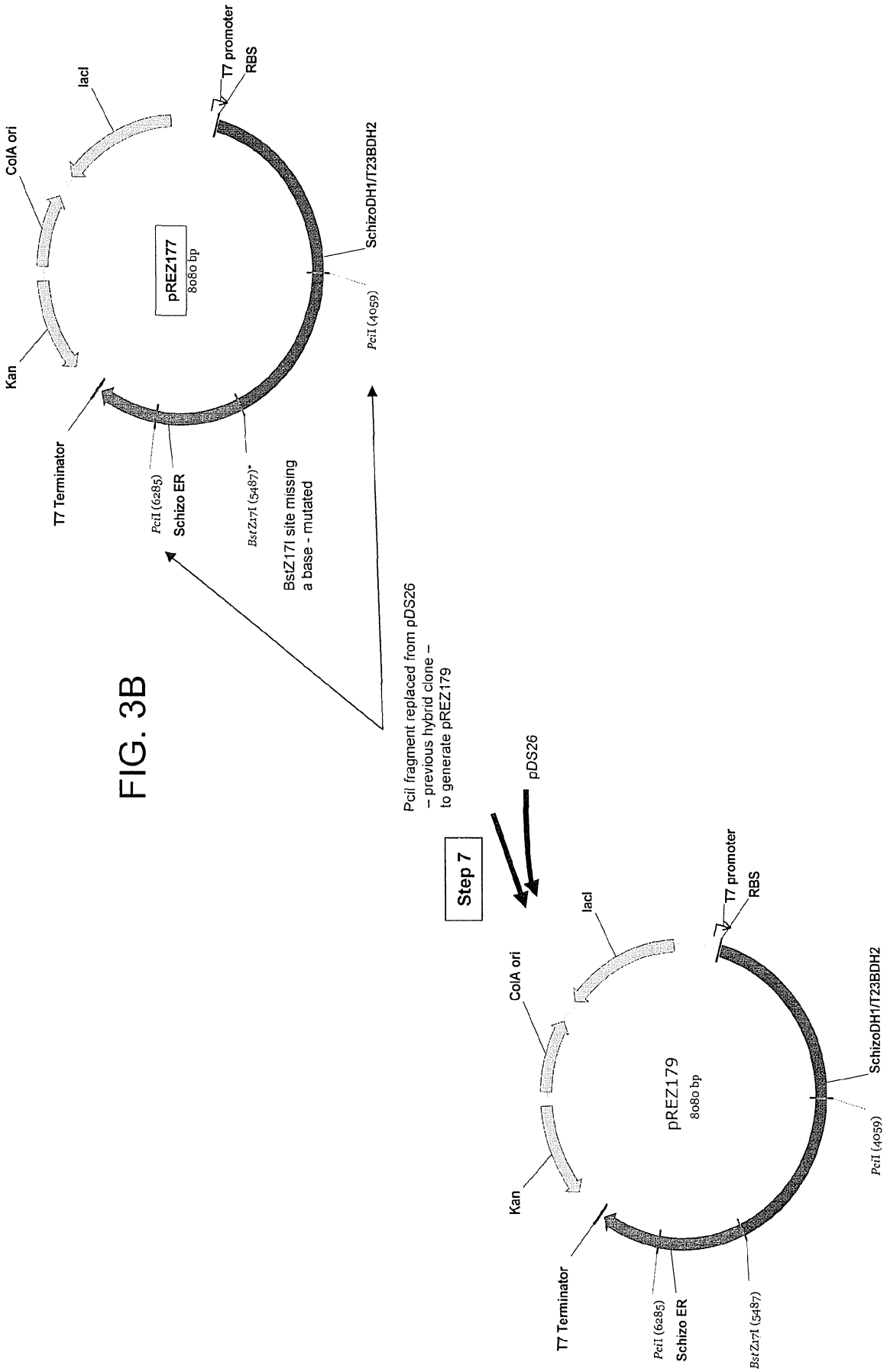

Step 7
pDS26
PciI fragment replaced from pDS26
– previous hybrid clone –
to generate pREZ179
pREZ179
8080 bp
ColA ori
lacI
T7 promoter
RBS
SchizoDH1/T23BDH2
PciI (4059)
Kan
T7 Terminator
PciI (6285)
Schizo ER
BstZ17I (5487)
pREZ177
8080 bp
BstZ17I site missing
a base - mutated

FIG. 3B


pREZ31
11312 bp
AmpR
ColE1 ori
lac promoter
flanking sequence
NheI (8006)
flanking sequence
BspEI (2661)
T23B orfC
BspEI (5931)
pBR002
12029 bp
Amp
Lac Promoter
NheI (9554)
Flanking Sequence
end
native Schizo orfC
BspEI (1914)
Flanking Sequence
BspEI (4126)
start
Step 8
pDS48
11395 bp
AmpR
ColE1 ori
lac promoter
flanking sequence
PflMI (7379)
native Schizo orfC
PstI (4129)
flanking sequence
Step 9
pREZ179
pDS49
11368 bp
AmpR
ColE1 ori
lac promoter
Flanking Sequence
Schizo
T23B
Schizo
flanking sequence

FIG. 3C

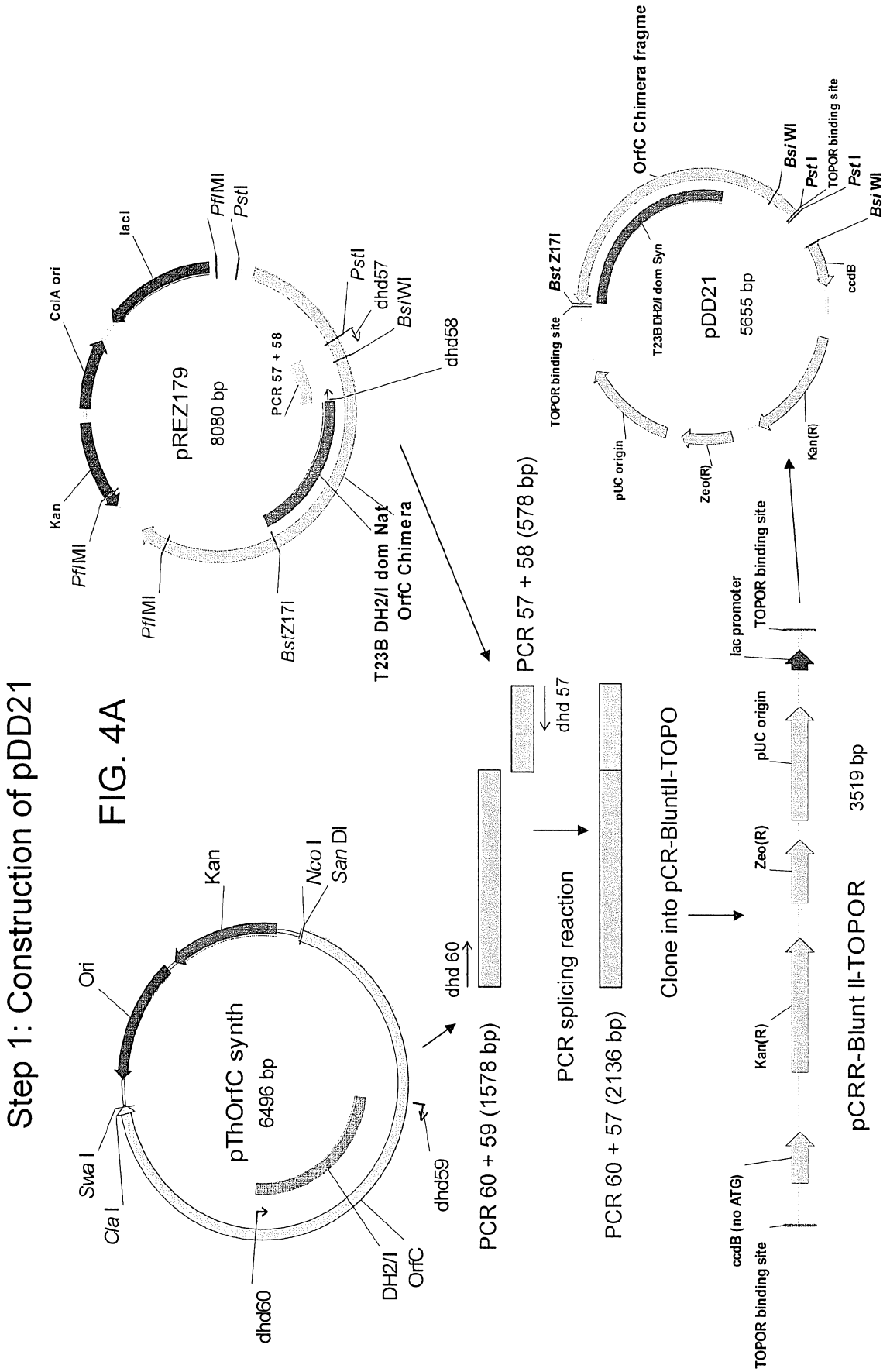
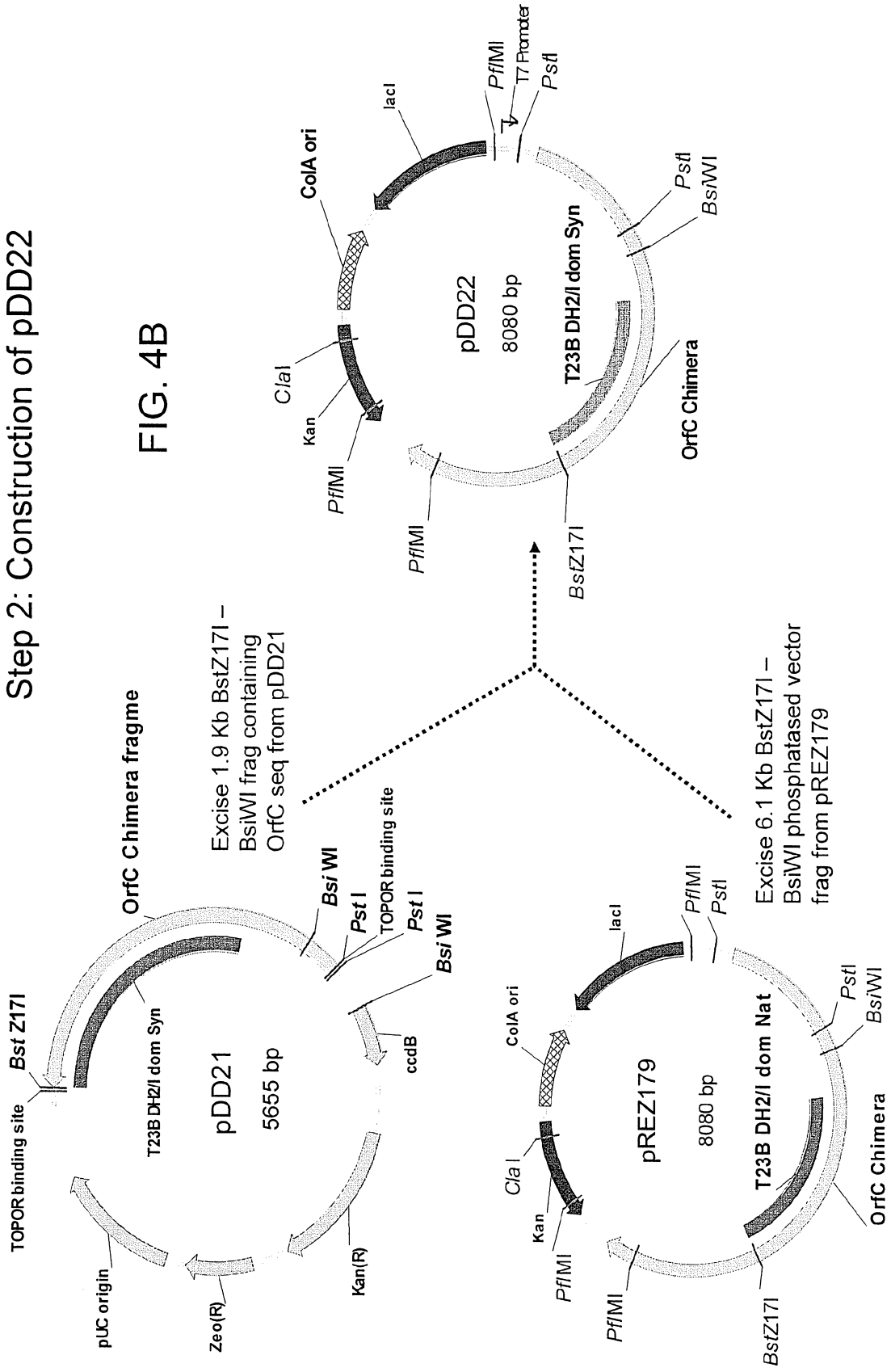

Step 1: Construction of pDD21
pREZ179
8080 bp
ColA ori
lacI
Kan
PflMI
PstI
PCR 57 + 58
dhd57
BsiWI
dhd58
BstZ17I
T23B DH2/I dom Nat
OrfC Chimera
pTh0rfC synth
6496 bp
Ori
Kan
Nco I
San DI
Swa I
Cla I
dhd60
DH2/I
OrfC
dhd59
dhd 60
PCR 60 + 59 (1578 bp)
PCR 57 + 58 (578 bp)
dhd 57
PCR splicing reaction
PCR 60 + 57 (2136 bp)
Clone into pCR-BluntII-TOPO
pDD21
5655 bp
OrfC Chimera fragme
Bst Z17I
T23B DH2/I dom Syn
TOPOR binding site
BsiWI
PstI
TOPOR binding site
PstI
BsiWI
ccdB
pUC origin
Zeo(R)
Kan(R)
lac promoter
TOPOR binding site
ccdB (no ATG)
TOPOR binding site
Kan(R)
Zeo(R)
pUC origin
pCRR-Blunt II-TOPOR
3519 bp

FIG. 4A


Step 2: Construction of pDD22
FIG. 4B
OrfC Chimera fragme
Bst Z17I
TOPOR binding site
T23B DH2/I dom Syn
pDD21
5655 bp
pUC origin
Zeo(R)
Kan(R)
ccdB
TOPOR binding site
Bsi WI
Pst I
Pst I
Bsi WI
Excise 1.9 Kb BstZ17I – BsiWI frag containing OrfC seq from pDD21
ColA ori
lacI
PflMI
T7 Promoter
PstI
ClaI
Kan
PflMI
pDD22
8080 bp
T23B DH2/I dom Syn
PstI
BsiWI
OrfC Chimera
PflMI
BstZ17I
Excise 6.1 Kb BstZ17I – BsiWI phosphatased vector frag from pREZ179
ColA ori
lacI
PflMI
PstI
ClaI
Kan
PflMI
pREZ179
8080 bp
T23B DH2/I dom Nat
PstI
BsiWI
OrfC Chimera
PflMI
BstZ17I

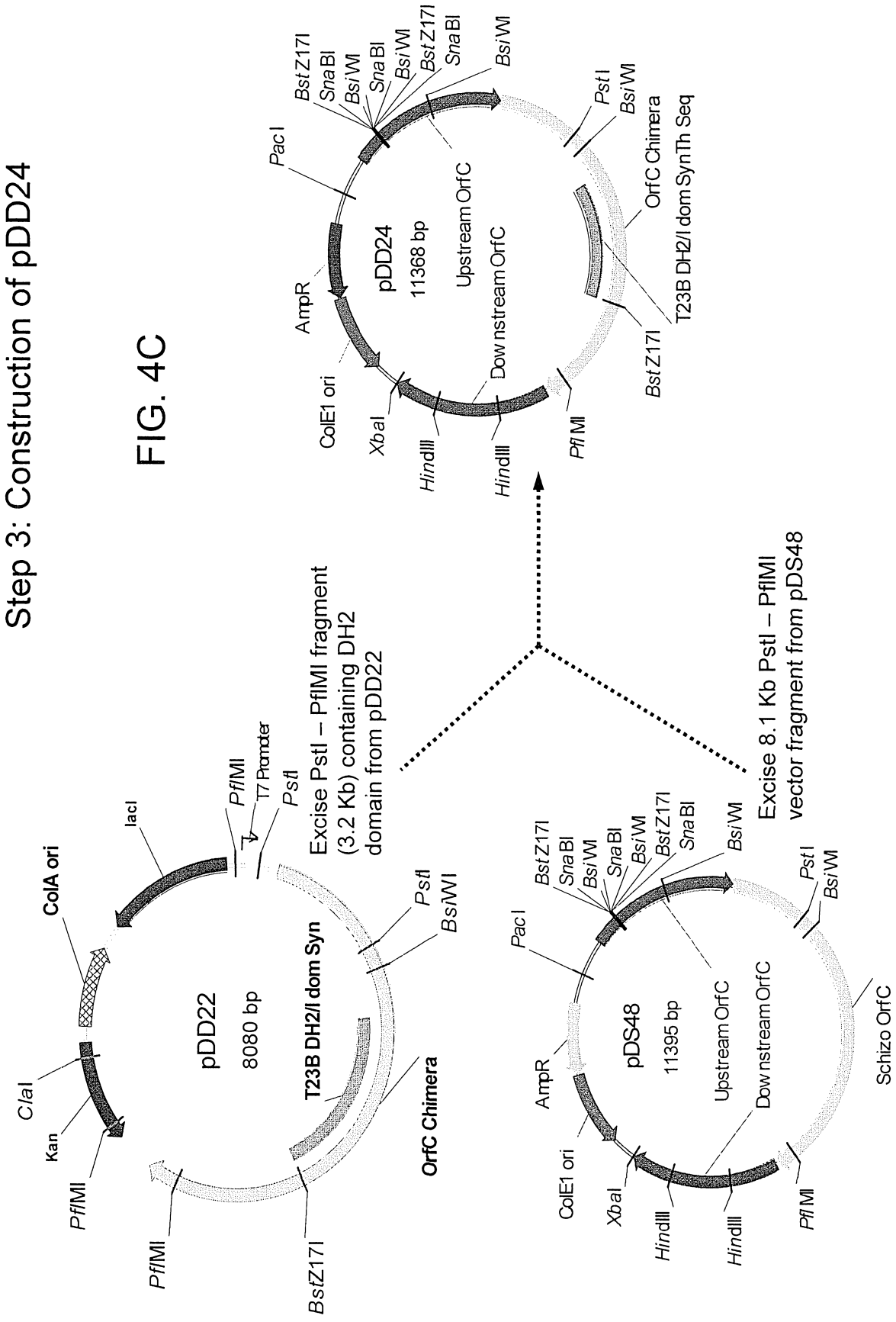

Step 3: Construction of pDD24
pDD22
8080 bp
ColA ori
lacI
PflMI
T7 Promoter
PstI
Excise PstI – PflMI fragment
(3.2 Kb) containing DH2
domain from pDD22
T23B DH2/I dom Syn
PstI
BsiWI
OrfC Chimera
ClaI
Kan
PflMI
PflMI
BstZ17I
pDD24
11368 bp
Upstream OrfC
Downstream OrfC
AmpR
ColE1 ori
XbaI
HindIII
HindIII
PflMI
BstZ17I
PacI
BstZ17I
SnaBI
BsiWI
SnaBI
BsiWI
BstZ17I
SnaBI
BsiWI
PstI
BsiWI
OrfC Chimera
T23B DH2/I dom SynTh Seq
pDS48
11395 bp
Upstream OrfC
Downstream OrfC
Excise 8.1 Kb PstI – PflMI
vector fragment from pDS48
Schizo OrfC

FIG. 4C

FIG. 5

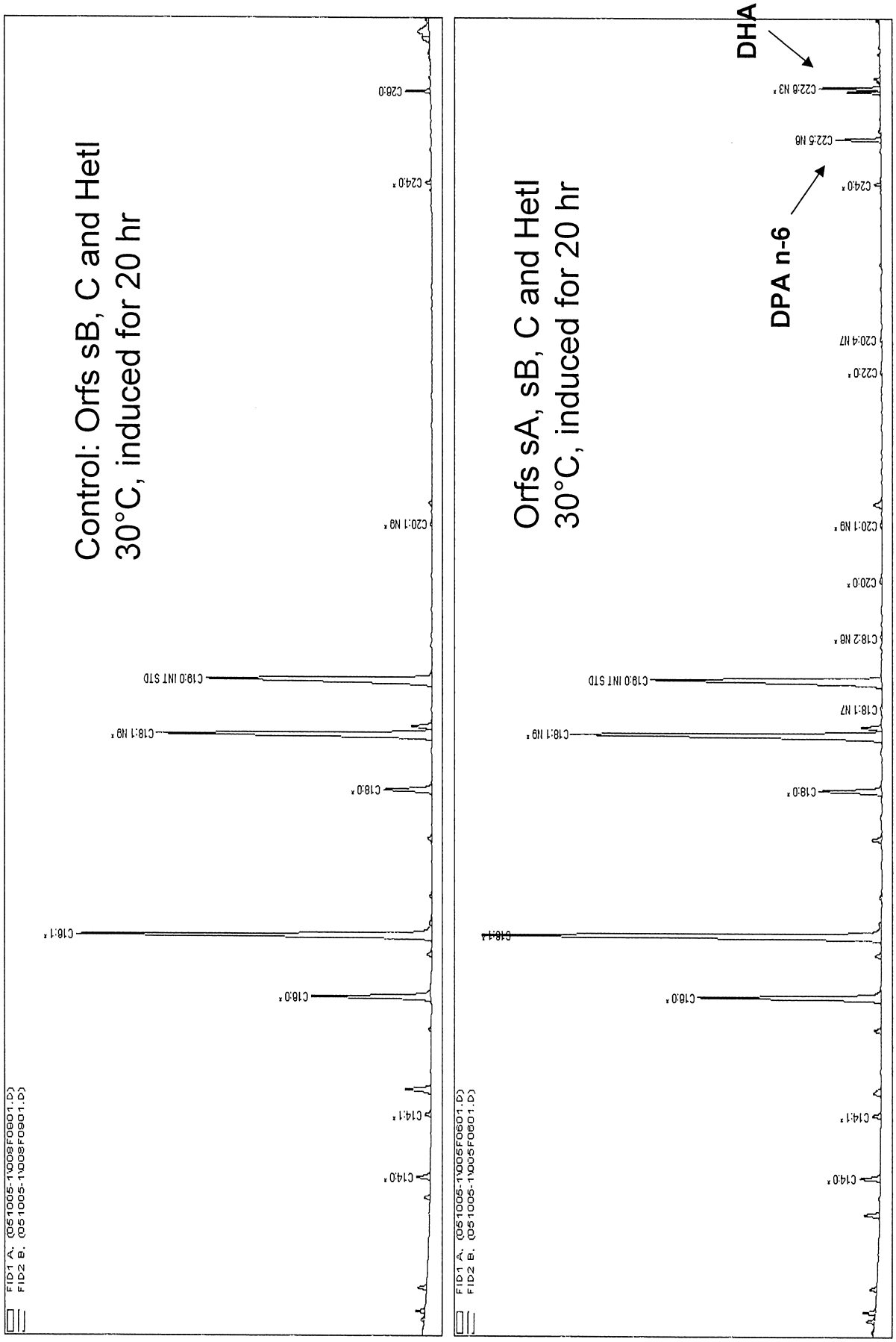
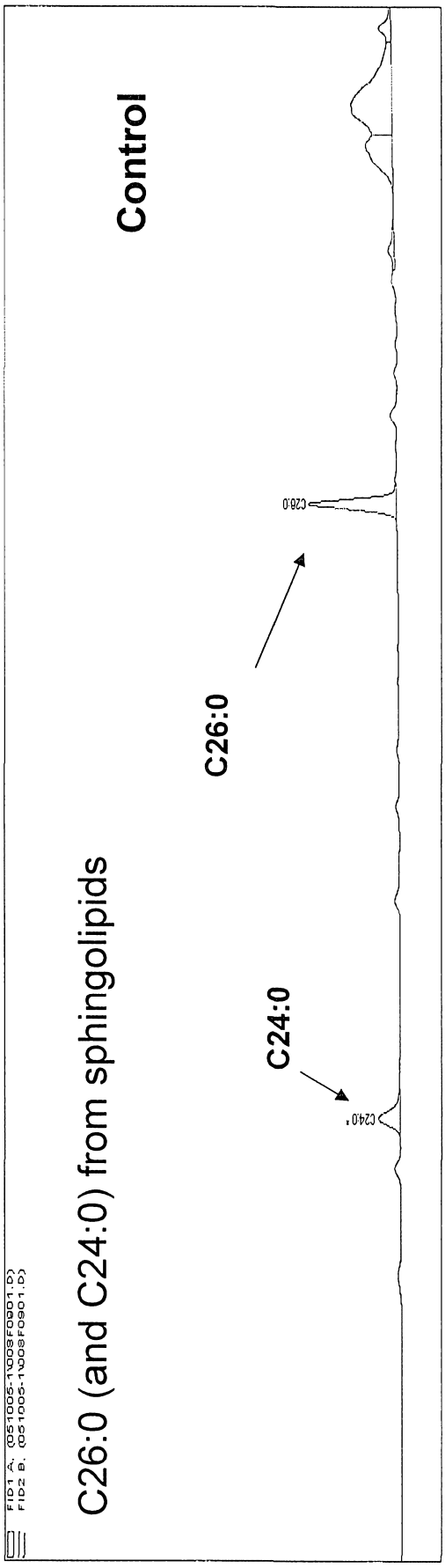
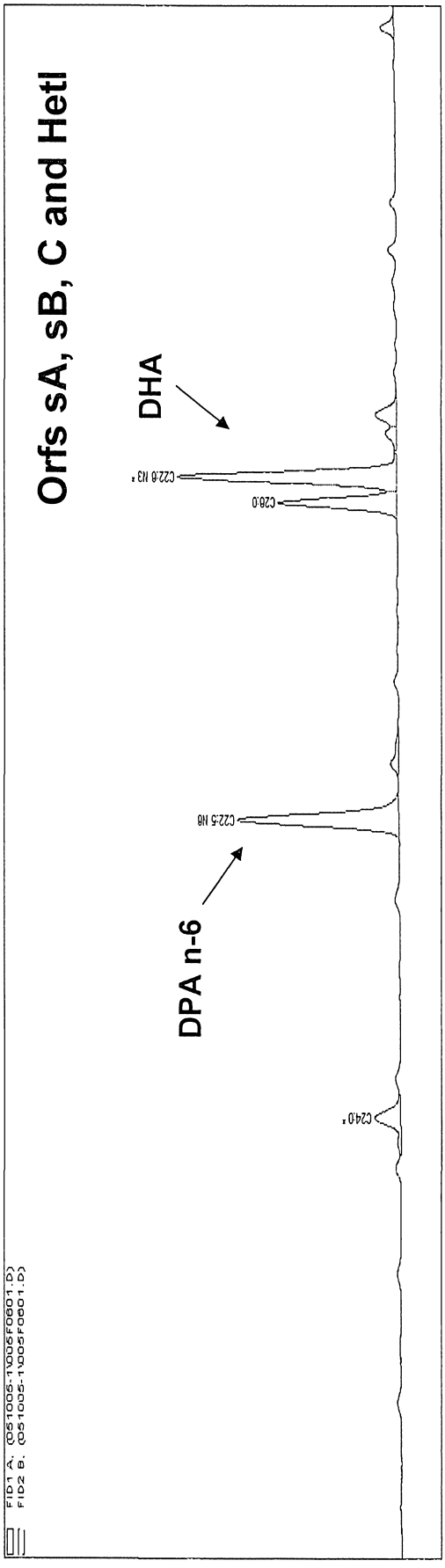

Control: Orfs sB, C and HetI
30°C, induced for 20 hr
Orfs sA, sB, C and HetI
30°C, induced for 20 hr
DHA
DPA n-6
C14:0
C14:1
C16:0
C16:1
C18:0
C18:1 N9
C19:0 INT STD
C20:1 N9
C24:0
C26:0
C18:1 N7
C18:2 N6
C20:0
C20:4 N7
C22:0
C22:5 N6
C22:6 N3
FID1 A, (051005-1\008F0901.D)
FID2 B, (051005-1\008F0901.D)
FID1 A, (051005-1\005F0601.D)
FID2 B, (051005-1\005F0901.D)

FIG. 6


Control
C26:0 (and C24:0) from sphingolipids
C26:0
C24:0
C26:0
C24:0
FID1 A, Q051006-1\006F0801.D)
FID2 B, Q051006-1\006F0801.D)


Orfs sA, sB, C and HetI
DHA
DPA n-6
C22:6 N3
C26:0
C22:5 N6
C24:0
FID1 A, Q051006-1\006F0801.D)
FID2 B, Q051006-1\006F0801.D)

SEQUENCE LISTING

<110> Martek Biosciences Corporation

<120> CHIMERIC PUFA POLYKETIDE SYNTHASE SYSTEMS

<130> 2997-49-2-PCT

<150> 11/749,686
<151> 2007-05-16

<160> 95

<170> PatentIn version 3.4

<210> 1
<211> 8733
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(8733)

<400> 1

```
atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc       48
Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
1               5                  10                  15

att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg       96
Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
            20                  25                  30

cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat      144
Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
        35                  40                  45

ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag      192
Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
    50                  55                  60

acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag      240
Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65                  70                  75                  80

tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag      288
Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
                85                  90                  95

gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc      336
Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
            100                 105                 110

ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc      384
Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
        115                 120                 125

ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc      432
Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
    130                 135                 140

tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg      480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145                 150                 155                 160

ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc      528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
```

```
                165                 170                 175

aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac      576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
            180                 185                 190

gtc acc gcc ggt cgc tgc acc aac acc ttc aac ctc gac ggc atg aac      624
Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
        195                 200                 205

tgc gtt gtc gac gcc gca tgc gcc tcg tcc ctc atc gcc gtc aag gtc      672
Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
    210                 215                 220

gcc atc gac gag ctg ctc tac ggt gac tgc gac atg atg gtc acc ggt      720
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
225                 230                 235                 240

gcc acc tgc acg gat aac tcc atc ggc atg tac atg gcc ttc tcc aag      768
Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
                245                 250                 255

acc ccc gtg ttc tcc acg gac ccc agc gtg cgc gcc tac gac gaa aag      816
Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
            260                 265                 270

aca aag ggc atg ctc atc ggc gag ggc tcc gcc atg ctc gtc ctc aag      864
Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
        275                 280                 285

cgc tac gcc gac gcc gtc cgc gac ggc gat gag atc cac gct gtt att      912
Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
    290                 295                 300

cgc ggc tgc gcc tcc tcc agt gat ggc aag gcc gcc ggc atc tac acg      960
Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
305                 310                 315                 320

ccc acc att tcg ggc cag gag gag gcc ctc cgc cgc gcc tac aac cgc     1008
Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
                325                 330                 335

gcc tgt gtc gac ccg gcc acc gtc act ctc gtc gag ggt cac ggc acc     1056
Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
            340                 345                 350

ggt act ccc gtt ggc gac cgc atc gag ctc acc gcc ttg cgc aac ctc     1104
Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
        355                 360                 365

ttt gac aag gcc tac ggc gag ggc aac acc gaa aag gtc gct gtg ggc     1152
Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
    370                 375                 380

agc atc aag tcc agc atc ggc cat ctc aag gcc gtc gcc ggt ctc gcc     1200
Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385                 390                 395                 400

ggt atg atc aag gtc atc atg gcg ctc aag cac aag act ctc ccg ggc     1248
Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
                405                 410                 415

acc atc aac gtc gac aac cca ccc aac ctc tac gac aac acg ccc atc     1296
Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
            420                 425                 430

aac gag tcc tcg ctc tac att aac acc atg aac cgc ccc tgg ttc ccg     1344
Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
```

```
        435                 440                 445

ccc cct ggt gtg ccc cgc cgc gcc ggc att tcg agc ttt ggc ttt ggt     1392
Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
    450                 455                 460

ggc gcc aac tac cac gcc gtc ctc gag gag gcc gag ccc gag cac acg     1440
Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465                 470                 475                 480

acc gcg tac cgc ctc aac aag cgc ccg cag ccc gtg ctc atg atg gcc     1488
Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
                485                 490                 495

gcc acg ccc gcg gcc ctc cag tcg ctc tgc gag gcc cag ctc aag gag     1536
Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu
            500                 505                 510

ttc gag gcc gcc atc aag gag aac gag acc gtc aag aac acc gcc tac     1584
Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr
        515                 520                 525

atc aag tgc gtc aag ttc ggc gag cag ttc aaa ttc cct ggc tcc atc     1632
Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile
    530                 535                 540

ccg gcc aca aac gcg cgc ctc ggc ttc ctc gtc aag gat gct gag gat     1680
Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp
545                 550                 555                 560

gcc tgc tcc acc ctc cgt gcc atc tgc gcc caa ttc gcc aag gat gtc     1728
Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val
                565                 570                 575

acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc cgc gcc     1776
Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala
            580                 585                 590

aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc ggc cag     1824
Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln
        595                 600                 605

ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac tgg ccc     1872
Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro
    610                 615                 620

cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc aag gtc     1920
Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val
625                 630                 635                 640

gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc tac ccg     1968
Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro
                645                 650                 655

cgc aag ccg tac gag cgt gag ccc gag cag gac cac aag aag atc tcc     2016
Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His Lys Lys Ile Ser
            660                 665                 670

ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc ggt gcc     2064
Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala
        675                 680                 685

ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc gcc ggc     2112
Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly
    690                 695                 700

cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc gtc gac     2160
His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp
```

```
705                 710                 715                 720

cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc atg ggc     2208
Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly
                725                 730                 735

ggc aag gac gca ccg gcc acc ccc aag ggc tgc atg gcc gcc gtc att     2256
Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile
            740                 745                 750

ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc tgg ctc     2304
Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu
        755                 760                 765

ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc gtc gaa     2352
Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu
    770                 775                 780

ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc cgc gtc     2400
Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val
785                 790                 795                 800

gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg gag aac     2448
Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn
                805                 810                 815

gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc cgc acc     2496
Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr
            820                 825                 830

ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag acc tac     2544
Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr
        835                 840                 845

ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc agc gtc     2592
Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val
    850                 855                 860

aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg cgc atc     2640
Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile
865                 870                 875                 880

ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc tcc gag     2688
Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu
                885                 890                 895

acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac ccg gcc     2736
Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala
            900                 905                 910

tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc cag ctc     2784
Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu
        915                 920                 925

gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac gcc ccc     2832
Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro
    930                 935                 940

gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc ctc cgc     2880
Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg
945                 950                 955                 960

ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc cgc gac     2928
Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp
                965                 970                 975

gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc gcc gca     2976
Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala
```

```
            980                 985                 990

ccg ctc atc aag gcc ccg gag ccc  gtt gtc gac gag gcc  gcc aag cgc   3024
Pro Leu Ile Lys Ala Pro Glu Pro  Val Val Asp Glu Ala  Ala Lys Arg
        995                1000                1005

gag gcc  gag cgt ctc cag aag  gag ctt cag gat gcc  cag cgc cag      3069
Glu Ala  Glu Arg Leu Gln Lys  Glu Leu Gln Asp Ala  Gln Arg Gln
    1010                1015                1020

ctc gac  gac gcc aag cgc gcc  gcc gcc gag gcc aac  tcc aag ctc      3114
Leu Asp  Asp Ala Lys Arg Ala  Ala Ala Glu Ala Asn  Ser Lys Leu
    1025                1030                1035

gcc gct  gcc aag gag gag gcc  aag acc gcc gct gct  tcg gcc aag      3159
Ala Ala  Ala Lys Glu Glu Ala  Lys Thr Ala Ala Ala  Ser Ala Lys
    1040                1045                1050

ccc gca  gtt gac act gct gtt  gtc gaa aag cat cgt  gcc atc ctc      3204
Pro Ala  Val Asp Thr Ala Val  Val Glu Lys His Arg  Ala Ile Leu
    1055                1060                1065

aag tcc  atg ctc gcg gag ctc  gat ggc tac gga tcg  gtc gac gct      3249
Lys Ser  Met Leu Ala Glu Leu  Asp Gly Tyr Gly Ser  Val Asp Ala
    1070                1075                1080

tct tcc  ctc cag cag cag cag  cag cag cag acg gcc  ccc gcc ccg      3294
Ser Ser  Leu Gln Gln Gln Gln  Gln Gln Gln Thr Ala  Pro Ala Pro
    1085                1090                1095

gtc aag  gct gct gcg cct gcc  gcc ccc gtt gcc tcg  gcc cct gcc      3339
Val Lys  Ala Ala Ala Pro Ala  Ala Pro Val Ala Ser  Ala Pro Ala
    1100                1105                1110

ccg gct  gtc tcg aac gag ctt  ctt gag aag gcc gag  act gtc gtc      3384
Pro Ala  Val Ser Asn Glu Leu  Leu Glu Lys Ala Glu  Thr Val Val
    1115                1120                1125

atg gag  gtc ctc gcc gcc aag  acc ggc tac gag acc  gac atg atc      3429
Met Glu  Val Leu Ala Ala Lys  Thr Gly Tyr Glu Thr  Asp Met Ile
    1130                1135                1140

gag gct  gac atg gag ctc gag  acc gag ctc ggc att  gac tcc atc      3474
Glu Ala  Asp Met Glu Leu Glu  Thr Glu Leu Gly Ile  Asp Ser Ile
    1145                1150                1155

aag cgt  gtc gag atc ctc tcc  gag gtc cag gcc atg  ctc aat gtc      3519
Lys Arg  Val Glu Ile Leu Ser  Glu Val Gln Ala Met  Leu Asn Val
    1160                1165                1170

gag gcc  aag gat gtc gat gcc  ctc agc cgc act cgc  act gtt ggt      3564
Glu Ala  Lys Asp Val Asp Ala  Leu Ser Arg Thr Arg  Thr Val Gly
    1175                1180                1185

gag gtt  gtc aac gcc atg aag  gcc gag atc gct ggc  agc tct gcc      3609
Glu Val  Val Asn Ala Met Lys  Ala Glu Ile Ala Gly  Ser Ser Ala
    1190                1195                1200

ccg gcg  cct gct gcc gct gct  ccg gct ccg gcc aag  gct gcc cct      3654
Pro Ala  Pro Ala Ala Ala Ala  Pro Ala Pro Ala Lys  Ala Ala Pro
    1205                1210                1215

gcc gcc  gct gcg cct gct gtc  tcg aac gag ctt ctc  gag aag gcc      3699
Ala Ala  Ala Ala Pro Ala Val  Ser Asn Glu Leu Leu  Glu Lys Ala
    1220                1225                1230

gag acc  gtc gtc atg gag gtc  ctc gcc gcc aag act  ggc tac gag      3744
Glu Thr  Val Val Met Glu Val  Leu Ala Ala Lys Thr  Gly Tyr Glu
```

```
    1235                1240                1245

act gac  atg atc gag tcc gac  atg gag ctc gag act  gag ctc ggc      3789
Thr Asp  Met Ile Glu Ser Asp  Met Glu Leu Glu Thr  Glu Leu Gly
    1250                1255                1260

att gac  tcc atc aag cgt gtc  gag atc ctc tcc gag  gtt cag gcc      3834
Ile Asp  Ser Ile Lys Arg Val  Glu Ile Leu Ser Glu  Val Gln Ala
    1265                1270                1275

atg ctc  aac gtc gag gcc aag  gac gtc gac gct ctc  agc cgc act      3879
Met Leu  Asn Val Glu Ala Lys  Asp Val Asp Ala Leu  Ser Arg Thr
    1280                1285                1290

cgc act  gtg ggt gag gtc gtc  aac gcc atg aag gct  gag atc gct      3924
Arg Thr  Val Gly Glu Val Val  Asn Ala Met Lys Ala  Glu Ile Ala
    1295                1300                1305

ggt ggc  tct gcc ccg gcg cct  gcc gcc gct gcc cca  ggt ccg gct      3969
Gly Gly  Ser Ala Pro Ala Pro  Ala Ala Ala Ala Pro  Gly Pro Ala
    1310                1315                1320

gct gcc  gcc cct gcg cct gcc  gcc gcc gcc cct gct  gtc tcg aac      4014
Ala Ala  Ala Pro Ala Pro Ala  Ala Ala Ala Pro Ala  Val Ser Asn
    1325                1330                1335

gag ctt  ctt gag aag gcc gag  acc gtc gtc atg gag  gtc ctc gcc      4059
Glu Leu  Leu Glu Lys Ala Glu  Thr Val Val Met Glu  Val Leu Ala
    1340                1345                1350

gcc aag  act ggc tac gag act  gac atg atc gag tcc  gac atg gag      4104
Ala Lys  Thr Gly Tyr Glu Thr  Asp Met Ile Glu Ser  Asp Met Glu
    1355                1360                1365

ctc gag  acc gag ctc ggc att  gac tcc atc aag cgt  gtc gag att      4149
Leu Glu  Thr Glu Leu Gly Ile  Asp Ser Ile Lys Arg  Val Glu Ile
    1370                1375                1380

ctc tcc  gag gtc cag gcc atg  ctc aac gtc gag gcc  aag gac gtc      4194
Leu Ser  Glu Val Gln Ala Met  Leu Asn Val Glu Ala  Lys Asp Val
    1385                1390                1395

gac gct  ctc agc cgc acc cgc  act gtt ggc gag gtc  gtc gat gcc      4239
Asp Ala  Leu Ser Arg Thr Arg  Thr Val Gly Glu Val  Val Asp Ala
    1400                1405                1410

atg aag  gcc gag atc gct ggt  ggc tct gcc ccg gcg  cct gcc gcc      4284
Met Lys  Ala Glu Ile Ala Gly  Gly Ser Ala Pro Ala  Pro Ala Ala
    1415                1420                1425

gct gct  cct gct ccg gct gct  gcc gcc cct gcg cct  gcc gcc cct      4329
Ala Ala  Pro Ala Pro Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Pro
    1430                1435                1440

gcg cct  gct gtc tcg agc gag  ctt ctc gag aag gcc  gag act gtc      4374
Ala Pro  Ala Val Ser Ser Glu  Leu Leu Glu Lys Ala  Glu Thr Val
    1445                1450                1455

gtc atg  gag gtc ctc gcc gcc  aag act ggc tac gag  act gac atg      4419
Val Met  Glu Val Leu Ala Ala  Lys Thr Gly Tyr Glu  Thr Asp Met
    1460                1465                1470

atc gag  tcc gac atg gag ctc  gag acc gag ctc ggc  att gac tcc      4464
Ile Glu  Ser Asp Met Glu Leu  Glu Thr Glu Leu Gly  Ile Asp Ser
    1475                1480                1485

atc aag  cgt gtc gag att ctc  tcc gag gtc cag gcc  atg ctc aac      4509
Ile Lys  Arg Val Glu Ile Leu  Ser Glu Val Gln Ala  Met Leu Asn
```

```
    1490                1495                1500

gtc gag  gcc aag gac gtc gac  gct ctc agc cgc acc  cgc act gtt      4554
Val Glu  Ala Lys Asp Val Asp  Ala Leu Ser Arg Thr  Arg Thr Val
    1505                1510                1515

ggc gag  gtc gtc gat gcc atg  aag gcc gag atc gct  ggt ggc tct      4599
Gly Glu  Val Val Asp Ala Met  Lys Ala Glu Ile Ala  Gly Gly Ser
    1520                1525                1530

gcc ccg  gcg cct gcc gcc gct  gct cct gct ccg gct  gct gcc gcc      4644
Ala Pro  Ala Pro Ala Ala Ala  Ala Pro Ala Pro Ala  Ala Ala Ala
    1535                1540                1545

cct gcg  cct gcc gcc cct gcg  cct gcc gcc cct gcg  cct gct gtc      4689
Pro Ala  Pro Ala Ala Pro Ala  Pro Ala Ala Pro Ala  Pro Ala Val
    1550                1555                1560

tcg agc  gag ctt ctc gag aag  gcc gag act gtc gtc  atg gag gtc      4734
Ser Ser  Glu Leu Leu Glu Lys  Ala Glu Thr Val Val  Met Glu Val
    1565                1570                1575

ctc gcc  gcc aag act ggc tac  gag act gac atg att  gag tcc gac      4779
Leu Ala  Ala Lys Thr Gly Tyr  Glu Thr Asp Met Ile  Glu Ser Asp
    1580                1585                1590

atg gag  ctc gag acc gag ctc  ggc att gac tcc atc  aag cgt gtc      4824
Met Glu  Leu Glu Thr Glu Leu  Gly Ile Asp Ser Ile  Lys Arg Val
    1595                1600                1605

gag att  ctc tcc gag gtt cag  gcc atg ctc aac gtc  gag gcc aag      4869
Glu Ile  Leu Ser Glu Val Gln  Ala Met Leu Asn Val  Glu Ala Lys
    1610                1615                1620

gac gtc  gac gct ctc agc cgc  act cgc act gtt ggt  gag gtc gtc      4914
Asp Val  Asp Ala Leu Ser Arg  Thr Arg Thr Val Gly  Glu Val Val
    1625                1630                1635

gat gcc  atg aag gct gag atc  gct ggc agc tcc gcc  tcg gcg cct      4959
Asp Ala  Met Lys Ala Glu Ile  Ala Gly Ser Ser Ala  Ser Ala Pro
    1640                1645                1650

gcc gcc  gct gct cct gct ccg  gct gct gcc gct cct  gcg ccc gct      5004
Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Ala Ala Pro  Ala Pro Ala
    1655                1660                1665

gcc gcc  gcc cct gct gtc tcg  aac gag ctt ctc gag  aaa gcc gag      5049
Ala Ala  Ala Pro Ala Val Ser  Asn Glu Leu Leu Glu  Lys Ala Glu
    1670                1675                1680

act gtc  gtc atg gag gtc ctc  gcc gcc aag act ggc  tac gag act      5094
Thr Val  Val Met Glu Val Leu  Ala Ala Lys Thr Gly  Tyr Glu Thr
    1685                1690                1695

gac atg  atc gag tcc gac atg  gag ctc gag act gag  ctc ggc att      5139
Asp Met  Ile Glu Ser Asp Met  Glu Leu Glu Thr Glu  Leu Gly Ile
    1700                1705                1710

gac tcc  atc aag cgt gtc gag  atc ctc tcc gag gtt  cag gcc atg      5184
Asp Ser  Ile Lys Arg Val Glu  Ile Leu Ser Glu Val  Gln Ala Met
    1715                1720                1725

ctc aac  gtc gag gcc aag gac  gtc gat gcc ctc agc  cgc acc cgc      5229
Leu Asn  Val Glu Ala Lys Asp  Val Asp Ala Leu Ser  Arg Thr Arg
    1730                1735                1740

act gtt  ggc gag gtt gtc gat  gcc atg aag gcc gag  atc gct ggt      5274
Thr Val  Gly Glu Val Val Asp  Ala Met Lys Ala Glu  Ile Ala Gly
```

```
    1745                1750                1755

ggc tct  gcc ccg gcg cct gcc  gcc gct gcc cct gct  ccg gct gcc      5319
Gly Ser  Ala Pro Ala Pro Ala  Ala Ala Ala Pro Ala  Pro Ala Ala
    1760                1765                1770

gcc gcc  cct gct gtc tcg aac  gag ctt ctc gag aag  gcc gag act      5364
Ala Ala  Pro Ala Val Ser Asn  Glu Leu Leu Glu Lys  Ala Glu Thr
    1775                1780                1785

gtc gtc  atg gag gtc ctc gcc  gcc aag act ggc tac  gag acc gac      5409
Val Val  Met Glu Val Leu Ala  Ala Lys Thr Gly Tyr  Glu Thr Asp
    1790                1795                1800

atg atc  gag tcc gac atg gag  ctc gag acc gag ctc  ggc att gac      5454
Met Ile  Glu Ser Asp Met Glu  Leu Glu Thr Glu Leu  Gly Ile Asp
    1805                1810                1815

tcc atc  aag cgt gtc gag att  ctc tcc gag gtt cag  gcc atg ctc      5499
Ser Ile  Lys Arg Val Glu Ile  Leu Ser Glu Val Gln  Ala Met Leu
    1820                1825                1830

aac gtc  gag gcc aag gac gtc  gat gct ctc agc cgc  act cgc act      5544
Asn Val  Glu Ala Lys Asp Val  Asp Ala Leu Ser Arg  Thr Arg Thr
    1835                1840                1845

gtt ggc  gag gtc gtc gat gcc  atg aag gct gag atc  gcc ggc agc      5589
Val Gly  Glu Val Val Asp Ala  Met Lys Ala Glu Ile  Ala Gly Ser
    1850                1855                1860

tcc gcc  ccg gcg cct gcc gcc  gct gct cct gct ccg  gct gct gcc      5634
Ser Ala  Pro Ala Pro Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Ala
    1865                1870                1875

gct cct  gcg ccc gct gcc gct  gcc cct gct gtc tcg  agc gag ctt      5679
Ala Pro  Ala Pro Ala Ala Ala  Ala Pro Ala Val Ser  Ser Glu Leu
    1880                1885                1890

ctc gag  aag gcc gag acc gtc  gtc atg gag gtc ctc  gcc gcc aag      5724
Leu Glu  Lys Ala Glu Thr Val  Val Met Glu Val Leu  Ala Ala Lys
    1895                1900                1905

act ggc  tac gag act gac atg  att gag tcc gac atg  gag ctc gag      5769
Thr Gly  Tyr Glu Thr Asp Met  Ile Glu Ser Asp Met  Glu Leu Glu
    1910                1915                1920

act gag  ctc ggc att gac tcc  atc aag cgt gtc gag  atc ctc tcc      5814
Thr Glu  Leu Gly Ile Asp Ser  Ile Lys Arg Val Glu  Ile Leu Ser
    1925                1930                1935

gag gtt  cag gcc atg ctc aac  gtc gag gcc aag gac  gtc gat gcc      5859
Glu Val  Gln Ala Met Leu Asn  Val Glu Ala Lys Asp  Val Asp Ala
    1940                1945                1950

ctc agc  cgc acc cgc act gtt  ggc gag gtt gtc gat  gcc atg aag      5904
Leu Ser  Arg Thr Arg Thr Val  Gly Glu Val Val Asp  Ala Met Lys
    1955                1960                1965

gcc gag  atc gct ggt ggc tct  gcc ccg gcg cct gcc  gcc gct gcc      5949
Ala Glu  Ile Ala Gly Gly Ser  Ala Pro Ala Pro Ala  Ala Ala Ala
    1970                1975                1980

cct gct  ccg gct gcc gcc gcc  cct gct gtc tcg aac  gag ctt ctt      5994
Pro Ala  Pro Ala Ala Ala Ala  Pro Ala Val Ser Asn  Glu Leu Leu
    1985                1990                1995

gag aag  gcc gag acc gtc gtc  atg gag gtc ctc gcc  gcc aag act      6039
Glu Lys  Ala Glu Thr Val Val  Met Glu Val Leu Ala  Ala Lys Thr
```

```
    2000                2005                2010

ggc tac  gag acc gac atg atc  gag tcc gac atg gag  ctc gag acc      6084
Gly Tyr  Glu Thr Asp Met Ile  Glu Ser Asp Met Glu  Leu Glu Thr
    2015                2020                2025

gag ctc  ggc att gac tcc atc  aag cgt gtc gag att  ctc tcc gag      6129
Glu Leu  Gly Ile Asp Ser Ile  Lys Arg Val Glu Ile  Leu Ser Glu
    2030                2035                2040

gtt cag  gcc atg ctc aac gtc  gag gcc aag gac gtc  gac gct ctc      6174
Val Gln  Ala Met Leu Asn Val  Glu Ala Lys Asp Val  Asp Ala Leu
    2045                2050                2055

agc cgc  act cgc act gtt ggc  gag gtc gtc gat gcc  atg aag gct      6219
Ser Arg  Thr Arg Thr Val Gly  Glu Val Val Asp Ala  Met Lys Ala
    2060                2065                2070

gag atc  gct ggt ggc tct gcc  ccg gcg cct gcc gcc  gct gct cct      6264
Glu Ile  Ala Gly Gly Ser Ala  Pro Ala Pro Ala Ala  Ala Ala Pro
    2075                2080                2085

gcc tcg  gct ggc gcc gcg cct  gcg gtc aag att gac  tcg gtc cac      6309
Ala Ser  Ala Gly Ala Ala Pro  Ala Val Lys Ile Asp  Ser Val His
    2090                2095                2100

ggc gct  gac tgt gat gat ctt  tcc ctg atg cac gcc  aag gtg gtt      6354
Gly Ala  Asp Cys Asp Asp Leu  Ser Leu Met His Ala  Lys Val Val
    2105                2110                2115

gac atc  cgc cgc ccg gac gag  ctc atc ctg gag cgc  ccc gag aac      6399
Asp Ile  Arg Arg Pro Asp Glu  Leu Ile Leu Glu Arg  Pro Glu Asn
    2120                2125                2130

cgc ccc  gtt ctc gtt gtc gat  gac ggc agc gag ctc  acc ctc gcc      6444
Arg Pro  Val Leu Val Val Asp  Asp Gly Ser Glu Leu  Thr Leu Ala
    2135                2140                2145

ctg gtc  cgc gtc ctc ggc gcc  tgc gcc gtt gtc ctg  acc ttt gag      6489
Leu Val  Arg Val Leu Gly Ala  Cys Ala Val Val Leu  Thr Phe Glu
    2150                2155                2160

ggt ctc  cag ctc gct cag cgc  gct ggt gcc gct gcc  atc cgc cac      6534
Gly Leu  Gln Leu Ala Gln Arg  Ala Gly Ala Ala Ala  Ile Arg His
    2165                2170                2175

gtg ctc  gcc aag gat ctt tcc  gcg gag agc gcc gag  aag gcc atc      6579
Val Leu  Ala Lys Asp Leu Ser  Ala Glu Ser Ala Glu  Lys Ala Ile
    2180                2185                2190

aag gag  gcc gag cag cgc ttt  ggc gct ctc ggc ggc  ttc atc tcg      6624
Lys Glu  Ala Glu Gln Arg Phe  Gly Ala Leu Gly Gly  Phe Ile Ser
    2195                2200                2205

cag cag  gcg gag cgc ttc gag  ccc gcc gaa atc ctc  ggc ttc acg      6669
Gln Gln  Ala Glu Arg Phe Glu  Pro Ala Glu Ile Leu  Gly Phe Thr
    2210                2215                2220

ctc atg  tgc gcc aag ttc gcc  aag gct tcc ctc tgc  acg gct gtg      6714
Leu Met  Cys Ala Lys Phe Ala  Lys Ala Ser Leu Cys  Thr Ala Val
    2225                2230                2235

gct ggc  ggc cgc ccg gcc ttt  atc ggt gtg gcg cgc  ctt gac ggc      6759
Ala Gly  Gly Arg Pro Ala Phe  Ile Gly Val Ala Arg  Leu Asp Gly
    2240                2245                2250

cgc ctc  gga ttc act tcg cag  ggc act tct gac gcg  ctc aag cgt      6804
Arg Leu  Gly Phe Thr Ser Gln  Gly Thr Ser Asp Ala  Leu Lys Arg
```

```
    2255                2260                2265

gcc cag  cgt ggt gcc atc ttt  ggc ctc tgc aag acc  atc ggc ctc      6849
Ala Gln  Arg Gly Ala Ile Phe  Gly Leu Cys Lys Thr  Ile Gly Leu
    2270                2275                2280

gag tgg  tcc gag tct gac gtc  ttt tcc cgc ggc gtg  gac att gct      6894
Glu Trp  Ser Glu Ser Asp Val  Phe Ser Arg Gly Val  Asp Ile Ala
    2285                2290                2295

cag ggc  atg cac ccc gag gat  gcc gcc gtg gcg att  gtg cgc gag      6939
Gln Gly  Met His Pro Glu Asp  Ala Ala Val Ala Ile  Val Arg Glu
    2300                2305                2310

atg gcg  tgc gct gac att cgc  att cgc gag gtc ggc  att ggc gca      6984
Met Ala  Cys Ala Asp Ile Arg  Ile Arg Glu Val Gly  Ile Gly Ala
    2315                2320                2325

aac cag  cag cgc tgc acg atc  cgt gcc gcc aag ctc  gag acc ggc      7029
Asn Gln  Gln Arg Cys Thr Ile  Arg Ala Ala Lys Leu  Glu Thr Gly
    2330                2335                2340

aac ccg  cag cgc cag atc gcc  aag gac gac gtg ctg  ctc gtt tct      7074
Asn Pro  Gln Arg Gln Ile Ala  Lys Asp Asp Val Leu  Leu Val Ser
    2345                2350                2355

ggc ggc  gct cgc ggc atc acg  cct ctt tgc atc cgg  gag atc acg      7119
Gly Gly  Ala Arg Gly Ile Thr  Pro Leu Cys Ile Arg  Glu Ile Thr
    2360                2365                2370

cgc cag  atc gcg ggc ggc aag  tac att ctg ctt ggc  cgc agc aag      7164
Arg Gln  Ile Ala Gly Gly Lys  Tyr Ile Leu Leu Gly  Arg Ser Lys
    2375                2380                2385

gtc tct  gcg agc gaa ccg gca  tgg tgc gct ggc atc  act gac gag      7209
Val Ser  Ala Ser Glu Pro Ala  Trp Cys Ala Gly Ile  Thr Asp Glu
    2390                2395                2400

aag gct  gtg caa aag gct gct  acc cag gag ctc aag  cgc gcc ttt      7254
Lys Ala  Val Gln Lys Ala Ala  Thr Gln Glu Leu Lys  Arg Ala Phe
    2405                2410                2415

agc gct  ggc gag ggc ccc aag  ccc acg ccc cgc gct  gtc act aag      7299
Ser Ala  Gly Glu Gly Pro Lys  Pro Thr Pro Arg Ala  Val Thr Lys
    2420                2425                2430

ctt gtg  ggc tct gtt ctt ggc  gct cgc gag gtg cgc  agc tct att      7344
Leu Val  Gly Ser Val Leu Gly  Ala Arg Glu Val Arg  Ser Ser Ile
    2435                2440                2445

gct gcg  att gaa gcg ctc ggc  ggc aag gcc atc tac  tcg tcg tgc      7389
Ala Ala  Ile Glu Ala Leu Gly  Gly Lys Ala Ile Tyr  Ser Ser Cys
    2450                2455                2460

gac gtg  aac tct gcc gcc gac  gtg gcc aag gcc gtg  cgc gat gcc      7434
Asp Val  Asn Ser Ala Ala Asp  Val Ala Lys Ala Val  Arg Asp Ala
    2465                2470                2475

gag tcc  cag ctc ggt gcc cgc  gtc tcg ggc atc gtt  cat gcc tcg      7479
Glu Ser  Gln Leu Gly Ala Arg  Val Ser Gly Ile Val  His Ala Ser
    2480                2485                2490

ggc gtg  ctc cgc gac cgt ctc  atc gag aag aag ctc  ccc gac gag      7524
Gly Val  Leu Arg Asp Arg Leu  Ile Glu Lys Lys Leu  Pro Asp Glu
    2495                2500                2505

ttc gac  gcc gtc ttt ggc acc  aag gtc acc ggt ctc  gag aac ctc      7569
Phe Asp  Ala Val Phe Gly Thr  Lys Val Thr Gly Leu  Glu Asn Leu
```

```
    2510                2515                2520

ctc gcc  gcc gtc gac cgc gcc  aac ctc aag cac atg  gtc ctc ttc      7614
Leu Ala  Ala Val Asp Arg Ala  Asn Leu Lys His Met  Val Leu Phe
    2525                2530                2535

agc tcg  ctc gcc ggc ttc cac  ggc aac gtc ggc cag  tct gac tac      7659
Ser Ser  Leu Ala Gly Phe His  Gly Asn Val Gly Gln  Ser Asp Tyr
    2540                2545                2550

gcc atg  gcc aac gag gcc ctt  aac aag atg ggc ctc  gag ctc gcc      7704
Ala Met  Ala Asn Glu Ala Leu  Asn Lys Met Gly Leu  Glu Leu Ala
    2555                2560                2565

aag gac  gtc tcg gtc aag tcg  atc tgc ttc ggt ccc  tgg gac ggt      7749
Lys Asp  Val Ser Val Lys Ser  Ile Cys Phe Gly Pro  Trp Asp Gly
    2570                2575                2580

ggc atg  gtg acg ccg cag ctc  aag aag cag ttc cag  gag atg ggc      7794
Gly Met  Val Thr Pro Gln Leu  Lys Lys Gln Phe Gln  Glu Met Gly
    2585                2590                2595

gtg cag  atc atc ccc cgc gag  ggc ggc gct gat acc  gtg gcg cgc      7839
Val Gln  Ile Ile Pro Arg Glu  Gly Gly Ala Asp Thr  Val Ala Arg
    2600                2605                2610

atc gtg  ctc ggc tcc tcg ccg  gct gag atc ctt gtc  ggc aac tgg      7884
Ile Val  Leu Gly Ser Ser Pro  Ala Glu Ile Leu Val  Gly Asn Trp
    2615                2620                2625

cgc acc  ccg tcc aag aag gtc  ggc tcg gac acc atc  acc ctg cac      7929
Arg Thr  Pro Ser Lys Lys Val  Gly Ser Asp Thr Ile  Thr Leu His
    2630                2635                2640

cgc aag  att tcc gcc aag tcc  aac ccc ttc ctc gag  gac cac gtc      7974
Arg Lys  Ile Ser Ala Lys Ser  Asn Pro Phe Leu Glu  Asp His Val
    2645                2650                2655

atc cag  ggc cgc cgc gtg ctg  ccc atg acg ctg gcc  att ggc tcg      8019
Ile Gln  Gly Arg Arg Val Leu  Pro Met Thr Leu Ala  Ile Gly Ser
    2660                2665                2670

ctc gcg  gag acc tgc ctc ggc  ctc ttc ccc ggc tac  tcg ctc tgg      8064
Leu Ala  Glu Thr Cys Leu Gly  Leu Phe Pro Gly Tyr  Ser Leu Trp
    2675                2680                2685

gcc att  gac gac gcc cag ctc  ttc aag ggt gtc act  gtc gac ggc      8109
Ala Ile  Asp Asp Ala Gln Leu  Phe Lys Gly Val Thr  Val Asp Gly
    2690                2695                2700

gac gtc  aac tgc gag gtg acc  ctc acc ccg tcg acg  gcg ccc tcg      8154
Asp Val  Asn Cys Glu Val Thr  Leu Thr Pro Ser Thr  Ala Pro Ser
    2705                2710                2715

ggc cgc  gtc aac gtc cag gcc  acg ctc aag acc ttt  tcc agc ggc      8199
Gly Arg  Val Asn Val Gln Ala  Thr Leu Lys Thr Phe  Ser Ser Gly
    2720                2725                2730

aag ctg  gtc ccg gcc tac cgc  gcc gtc atc gtg ctc  tcc aac cag      8244
Lys Leu  Val Pro Ala Tyr Arg  Ala Val Ile Val Leu  Ser Asn Gln
    2735                2740                2745

ggc gcg  ccc ccg gcc aac gcc  acc atg cag ccg ccc  tcg ctc gat      8289
Gly Ala  Pro Pro Ala Asn Ala  Thr Met Gln Pro Pro  Ser Leu Asp
    2750                2755                2760

gcc gat  ccg gcg ctc cag ggc  tcc gtc tac gac ggc  aag acc ctc      8334
Ala Asp  Pro Ala Leu Gln Gly  Ser Val Tyr Asp Gly  Lys Thr Leu
```

```
    2765                2770                2775

ttc cac  ggc ccg gcc ttc cgc  ggc atc gat gac gtg  ctc tcg tgc      8379
Phe His  Gly Pro Ala Phe Arg  Gly Ile Asp Asp Val  Leu Ser Cys
    2780                2785                2790

acc aag  agc cag ctt gtg gcc  aag tgc agc gct gtc  ccc ggc tcc      8424
Thr Lys  Ser Gln Leu Val Ala  Lys Cys Ser Ala Val  Pro Gly Ser
    2795                2800                2805

gac gcc  gct cgc ggc gag ttt  gcc acg gac act gac  gcc cat gac      8469
Asp Ala  Ala Arg Gly Glu Phe  Ala Thr Asp Thr Asp  Ala His Asp
    2810                2815                2820

ccc ttc  gtg aac gac ctg gcc  ttt cag gcc atg ctc  gtc tgg gtg      8514
Pro Phe  Val Asn Asp Leu Ala  Phe Gln Ala Met Leu  Val Trp Val
    2825                2830                2835

cgc cgc  acg ctc ggc cag gct  gcg ctc ccc aac tcg  atc cag cgc      8559
Arg Arg  Thr Leu Gly Gln Ala  Ala Leu Pro Asn Ser  Ile Gln Arg
    2840                2845                2850

atc gtc  cag cac cgc ccg gtc  ccg cag gac aag ccc  ttc tac att      8604
Ile Val  Gln His Arg Pro Val  Pro Gln Asp Lys Pro  Phe Tyr Ile
    2855                2860                2865

acc ctc  cgc tcc aac cag tcg  ggc ggt cac tcc cag  cac aag cac      8649
Thr Leu  Arg Ser Asn Gln Ser  Gly Gly His Ser Gln  His Lys His
    2870                2875                2880

gcc ctt  cag ttc cac aac gag  cag ggc gat ctc ttc  att gat gtc      8694
Ala Leu  Gln Phe His Asn Glu  Gln Gly Asp Leu Phe  Ile Asp Val
    2885                2890                2895

cag gct  tcg gtc atc gcc acg  gac agc ctt gcc ttc  taa              8733
Gln Ala  Ser Val Ile Ala Thr  Asp Ser Leu Ala Phe
    2900                2905                2910
```

<210> 2
<211> 2910
<212> PRT
<213> Schizochytrium sp.

<400> 2

```
Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
1               5                   10                  15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
            20                  25                  30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
        35                  40                  45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
    50                  55                  60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65                  70                  75                  80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
                85                  90                  95
```

```
Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
            100                 105                 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
        115                 120                 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
    130                 135                 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145                 150                 155                 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
                165                 170                 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
            180                 185                 190

Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
        195                 200                 205

Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
    210                 215                 220

Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
225                 230                 235                 240

Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
                245                 250                 255

Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
            260                 265                 270

Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
        275                 280                 285

Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
    290                 295                 300

Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
305                 310                 315                 320

Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
                325                 330                 335

Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
            340                 345                 350

Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
        355                 360                 365
```

```
Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
    370                 375                 380

Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385                 390                 395                 400

Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
                405                 410                 415

Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
            420                 425                 430

Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
        435                 440                 445

Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
    450                 455                 460

Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465                 470                 475                 480

Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
                485                 490                 495

Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu
            500                 505                 510

Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr
        515                 520                 525

Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile
    530                 535                 540

Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp
545                 550                 555                 560

Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val
                565                 570                 575

Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala
            580                 585                 590

Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln
        595                 600                 605

Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro
    610                 615                 620

Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val
625                 630                 635                 640
```

```
Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro
                645                 650                 655

Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His Lys Lys Ile Ser
            660                 665                 670

Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala
        675                 680                 685

Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly
    690                 695                 700

His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp
705                 710                 715                 720

Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly
                725                 730                 735

Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile
            740                 745                 750

Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu
        755                 760                 765

Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu
    770                 775                 780

Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val
785                 790                 795                 800

Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn
                805                 810                 815

Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr
            820                 825                 830

Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr
        835                 840                 845

Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val
    850                 855                 860

Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile
865                 870                 875                 880

Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu
                885                 890                 895

Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala
            900                 905                 910
```

```
Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu
        915                 920                 925

Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro
    930                 935                 940

Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg
945                 950                 955                 960

Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp
                965                 970                 975

Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala
            980                 985                 990

Pro Leu Ile Lys Ala Pro Glu Pro  Val Val Asp Glu Ala  Ala Lys Arg
        995                 1000                1005

Glu Ala  Glu Arg Leu Gln Lys  Glu Leu Gln Asp Ala  Gln Arg Gln
    1010                1015                1020

Leu Asp  Asp Ala Lys Arg Ala  Ala Ala Glu Ala Asn  Ser Lys Leu
    1025                1030                1035

Ala Ala  Ala Lys Glu Glu Ala  Lys Thr Ala Ala Ala  Ser Ala Lys
    1040                1045                1050

Pro Ala  Val Asp Thr Ala Val  Val Glu Lys His Arg  Ala Ile Leu
    1055                1060                1065

Lys Ser  Met Leu Ala Glu Leu  Asp Gly Tyr Gly Ser  Val Asp Ala
    1070                1075                1080

Ser Ser  Leu Gln Gln Gln Gln  Gln Gln Gln Thr Ala  Pro Ala Pro
    1085                1090                1095

Val Lys  Ala Ala Ala Pro Ala  Ala Pro Val Ala Ser  Ala Pro Ala
    1100                1105                1110

Pro Ala  Val Ser Asn Glu Leu  Leu Glu Lys Ala Glu  Thr Val Val
    1115                1120                1125

Met Glu  Val Leu Ala Ala Lys  Thr Gly Tyr Glu Thr  Asp Met Ile
    1130                1135                1140

Glu Ala  Asp Met Glu Leu Glu  Thr Glu Leu Gly Ile  Asp Ser Ile
    1145                1150                1155

Lys Arg  Val Glu Ile Leu Ser  Glu Val Gln Ala Met  Leu Asn Val
    1160                1165                1170
```

```
Glu Ala  Lys Asp Val Asp Ala  Leu Ser Arg Thr Arg  Thr Val Gly
    1175                1180                1185

Glu Val  Val Asn Ala Met Lys  Ala Glu Ile Ala Gly  Ser Ser Ala
    1190                1195                1200

Pro Ala  Pro Ala Ala Ala Ala  Pro Ala Pro Ala Lys  Ala Ala Pro
    1205                1210                1215

Ala Ala  Ala Ala Pro Ala Val  Ser Asn Glu Leu Leu  Glu Lys Ala
    1220                1225                1230

Glu Thr  Val Val Met Glu Val  Leu Ala Ala Lys Thr  Gly Tyr Glu
    1235                1240                1245

Thr Asp  Met Ile Glu Ser Asp  Met Glu Leu Glu Thr  Glu Leu Gly
    1250                1255                1260

Ile Asp  Ser Ile Lys Arg Val  Glu Ile Leu Ser Glu  Val Gln Ala
    1265                1270                1275

Met Leu  Asn Val Glu Ala Lys  Asp Val Asp Ala Leu  Ser Arg Thr
    1280                1285                1290

Arg Thr  Val Gly Glu Val Val  Asn Ala Met Lys Ala  Glu Ile Ala
    1295                1300                1305

Gly Gly  Ser Ala Pro Ala Pro  Ala Ala Ala Ala Pro  Gly Pro Ala
    1310                1315                1320

Ala Ala  Ala Pro Ala Pro Ala  Ala Ala Ala Pro Ala  Val Ser Asn
    1325                1330                1335

Glu Leu  Leu Glu Lys Ala Glu  Thr Val Val Met Glu  Val Leu Ala
    1340                1345                1350

Ala Lys  Thr Gly Tyr Glu Thr  Asp Met Ile Glu Ser  Asp Met Glu
    1355                1360                1365

Leu Glu  Thr Glu Leu Gly Ile  Asp Ser Ile Lys Arg  Val Glu Ile
    1370                1375                1380

Leu Ser  Glu Val Gln Ala Met  Leu Asn Val Glu Ala  Lys Asp Val
    1385                1390                1395

Asp Ala  Leu Ser Arg Thr Arg  Thr Val Gly Glu Val  Val Asp Ala
    1400                1405                1410

Met Lys  Ala Glu Ile Ala Gly  Gly Ser Ala Pro Ala  Pro Ala Ala
    1415                1420                1425
```

```
Ala Ala  Pro Ala Pro Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Pro
    1430                 1435                 1440

Ala Pro  Ala Val Ser Ser Glu  Leu Leu Glu Lys Ala  Glu Thr Val
    1445                 1450                 1455

Val Met  Glu Val Leu Ala Ala  Lys Thr Gly Tyr Glu  Thr Asp Met
    1460                 1465                 1470

Ile Glu  Ser Asp Met Glu Leu  Glu Thr Glu Leu Gly  Ile Asp Ser
    1475                 1480                 1485

Ile Lys  Arg Val Glu Ile Leu  Ser Glu Val Gln Ala  Met Leu Asn
    1490                 1495                 1500

Val Glu  Ala Lys Asp Val Asp  Ala Leu Ser Arg Thr  Arg Thr Val
    1505                 1510                 1515

Gly Glu  Val Val Asp Ala Met  Lys Ala Glu Ile Ala  Gly Gly Ser
    1520                 1525                 1530

Ala Pro  Ala Pro Ala Ala Ala  Ala Pro Ala Pro Ala  Ala Ala Ala
    1535                 1540                 1545

Pro Ala  Pro Ala Ala Pro Ala  Pro Ala Ala Pro Ala  Pro Ala Val
    1550                 1555                 1560

Ser Ser  Glu Leu Leu Glu Lys  Ala Glu Thr Val Val  Met Glu Val
    1565                 1570                 1575

Leu Ala  Ala Lys Thr Gly Tyr  Glu Thr Asp Met Ile  Glu Ser Asp
    1580                 1585                 1590

Met Glu  Leu Glu Thr Glu Leu  Gly Ile Asp Ser Ile  Lys Arg Val
    1595                 1600                 1605

Glu Ile  Leu Ser Glu Val Gln  Ala Met Leu Asn Val  Glu Ala Lys
    1610                 1615                 1620

Asp Val  Asp Ala Leu Ser Arg  Thr Arg Thr Val Gly  Glu Val Val
    1625                 1630                 1635

Asp Ala  Met Lys Ala Glu Ile  Ala Gly Ser Ser Ala  Ser Ala Pro
    1640                 1645                 1650

Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Ala Ala Pro  Ala Pro Ala
    1655                 1660                 1665

Ala Ala  Ala Pro Ala Val Ser  Asn Glu Leu Leu Glu  Lys Ala Glu
    1670                 1675                 1680
```

```
Thr Val  Val Met Glu Val Leu  Ala Ala Lys Thr Gly  Tyr Glu Thr
    1685                 1690                 1695

Asp Met  Ile Glu Ser Asp Met  Glu Leu Glu Thr Glu  Leu Gly Ile
    1700                 1705                 1710

Asp Ser  Ile Lys Arg Val Glu  Ile Leu Ser Glu Val  Gln Ala Met
    1715                 1720                 1725

Leu Asn  Val Glu Ala Lys Asp  Val Asp Ala Leu Ser  Arg Thr Arg
    1730                 1735                 1740

Thr Val  Gly Glu Val Val Asp  Ala Met Lys Ala Glu  Ile Ala Gly
    1745                 1750                 1755

Gly Ser  Ala Pro Ala Pro Ala  Ala Ala Ala Pro Ala  Pro Ala Ala
    1760                 1765                 1770

Ala Ala  Pro Ala Val Ser Asn  Glu Leu Leu Glu Lys  Ala Glu Thr
    1775                 1780                 1785

Val Val  Met Glu Val Leu Ala  Ala Lys Thr Gly Tyr  Glu Thr Asp
    1790                 1795                 1800

Met Ile  Glu Ser Asp Met Glu  Leu Glu Thr Glu Leu  Gly Ile Asp
    1805                 1810                 1815

Ser Ile  Lys Arg Val Glu Ile  Leu Ser Glu Val Gln  Ala Met Leu
    1820                 1825                 1830

Asn Val  Glu Ala Lys Asp Val  Asp Ala Leu Ser Arg  Thr Arg Thr
    1835                 1840                 1845

Val Gly  Glu Val Val Asp Ala  Met Lys Ala Glu Ile  Ala Gly Ser
    1850                 1855                 1860

Ser Ala  Pro Ala Pro Ala Ala  Ala Ala Pro Ala Pro  Ala Ala Ala
    1865                 1870                 1875

Ala Pro  Ala Pro Ala Ala Ala  Ala Pro Ala Val Ser  Ser Glu Leu
    1880                 1885                 1890

Leu Glu  Lys Ala Glu Thr Val  Val Met Glu Val Leu  Ala Ala Lys
    1895                 1900                 1905

Thr Gly  Tyr Glu Thr Asp Met  Ile Glu Ser Asp Met  Glu Leu Glu
    1910                 1915                 1920

Thr Glu  Leu Gly Ile Asp Ser  Ile Lys Arg Val Glu  Ile Leu Ser
    1925                 1930                 1935
```

```
Glu Val  Gln Ala Met Leu Asn  Val Glu Ala Lys Asp  Val Asp Ala
    1940                 1945                 1950

Leu Ser  Arg Thr Arg Thr Val  Gly Glu Val Val Asp  Ala Met Lys
    1955                 1960                 1965

Ala Glu  Ile Ala Gly Gly Ser  Ala Pro Ala Pro Ala  Ala Ala Ala
    1970                 1975                 1980

Pro Ala  Pro Ala Ala Ala Ala  Pro Ala Val Ser Asn  Glu Leu Leu
    1985                 1990                 1995

Glu Lys  Ala Glu Thr Val Val  Met Glu Val Leu Ala  Ala Lys Thr
    2000                 2005                 2010

Gly Tyr  Glu Thr Asp Met Ile  Glu Ser Asp Met Glu  Leu Glu Thr
    2015                 2020                 2025

Glu Leu  Gly Ile Asp Ser Ile  Lys Arg Val Glu Ile  Leu Ser Glu
    2030                 2035                 2040

Val Gln  Ala Met Leu Asn Val  Glu Ala Lys Asp Val  Asp Ala Leu
    2045                 2050                 2055

Ser Arg  Thr Arg Thr Val Gly  Glu Val Val Asp Ala  Met Lys Ala
    2060                 2065                 2070

Glu Ile  Ala Gly Gly Ser Ala  Pro Ala Pro Ala Ala  Ala Ala Pro
    2075                 2080                 2085

Ala Ser  Ala Gly Ala Ala Pro  Ala Val Lys Ile Asp  Ser Val His
    2090                 2095                 2100

Gly Ala  Asp Cys Asp Asp Leu  Ser Leu Met His Ala  Lys Val Val
    2105                 2110                 2115

Asp Ile  Arg Arg Pro Asp Glu  Leu Ile Leu Glu Arg  Pro Glu Asn
    2120                 2125                 2130

Arg Pro  Val Leu Val Val Asp  Asp Gly Ser Glu Leu  Thr Leu Ala
    2135                 2140                 2145

Leu Val  Arg Val Leu Gly Ala  Cys Ala Val Val Leu  Thr Phe Glu
    2150                 2155                 2160

Gly Leu  Gln Leu Ala Gln Arg  Ala Gly Ala Ala Ala  Ile Arg His
    2165                 2170                 2175

Val Leu  Ala Lys Asp Leu Ser  Ala Glu Ser Ala Glu  Lys Ala Ile
    2180                 2185                 2190
```

```
Lys Glu  Ala Glu Gln Arg Phe  Gly Ala Leu Gly Gly  Phe Ile Ser
    2195                 2200                 2205

Gln Gln  Ala Glu Arg Phe Glu  Pro Ala Glu Ile Leu  Gly Phe Thr
    2210                 2215                 2220

Leu Met  Cys Ala Lys Phe Ala  Lys Ala Ser Leu Cys  Thr Ala Val
    2225                 2230                 2235

Ala Gly  Gly Arg Pro Ala Phe  Ile Gly Val Ala Arg  Leu Asp Gly
    2240                 2245                 2250

Arg Leu  Gly Phe Thr Ser Gln  Gly Thr Ser Asp Ala  Leu Lys Arg
    2255                 2260                 2265

Ala Gln  Arg Gly Ala Ile Phe  Gly Leu Cys Lys Thr  Ile Gly Leu
    2270                 2275                 2280

Glu Trp  Ser Glu Ser Asp Val  Phe Ser Arg Gly Val  Asp Ile Ala
    2285                 2290                 2295

Gln Gly  Met His Pro Glu Asp  Ala Ala Val Ala Ile  Val Arg Glu
    2300                 2305                 2310

Met Ala  Cys Ala Asp Ile Arg  Ile Arg Glu Val Gly  Ile Gly Ala
    2315                 2320                 2325

Asn Gln  Gln Arg Cys Thr Ile  Arg Ala Ala Lys Leu  Glu Thr Gly
    2330                 2335                 2340

Asn Pro  Gln Arg Gln Ile Ala  Lys Asp Asp Val Leu  Leu Val Ser
    2345                 2350                 2355

Gly Gly  Ala Arg Gly Ile Thr  Pro Leu Cys Ile Arg  Glu Ile Thr
    2360                 2365                 2370

Arg Gln  Ile Ala Gly Gly Lys  Tyr Ile Leu Leu Gly  Arg Ser Lys
    2375                 2380                 2385

Val Ser  Ala Ser Glu Pro Ala  Trp Cys Ala Gly Ile  Thr Asp Glu
    2390                 2395                 2400

Lys Ala  Val Gln Lys Ala Ala  Thr Gln Glu Leu Lys  Arg Ala Phe
    2405                 2410                 2415

Ser Ala  Gly Glu Gly Pro Lys  Pro Thr Pro Arg Ala  Val Thr Lys
    2420                 2425                 2430

Leu Val  Gly Ser Val Leu Gly  Ala Arg Glu Val Arg  Ser Ser Ile
    2435                 2440                 2445
```

```
Ala Ala  Ile Glu Ala Leu Gly  Gly Lys Ala Ile Tyr  Ser Ser Cys
    2450                 2455                 2460

Asp Val  Asn Ser Ala Ala Asp  Val Ala Lys Ala Val  Arg Asp Ala
    2465                 2470                 2475

Glu Ser  Gln Leu Gly Ala Arg  Val Ser Gly Ile Val  His Ala Ser
    2480                 2485                 2490

Gly Val  Leu Arg Asp Arg Leu  Ile Glu Lys Lys Leu  Pro Asp Glu
    2495                 2500                 2505

Phe Asp  Ala Val Phe Gly Thr  Lys Val Thr Gly Leu  Glu Asn Leu
    2510                 2515                 2520

Leu Ala  Ala Val Asp Arg Ala  Asn Leu Lys His Met  Val Leu Phe
    2525                 2530                 2535

Ser Ser  Leu Ala Gly Phe His  Gly Asn Val Gly Gln  Ser Asp Tyr
    2540                 2545                 2550

Ala Met  Ala Asn Glu Ala Leu  Asn Lys Met Gly Leu  Glu Leu Ala
    2555                 2560                 2565

Lys Asp  Val Ser Val Lys Ser  Ile Cys Phe Gly Pro  Trp Asp Gly
    2570                 2575                 2580

Gly Met  Val Thr Pro Gln Leu  Lys Lys Gln Phe Gln  Glu Met Gly
    2585                 2590                 2595

Val Gln  Ile Ile Pro Arg Glu  Gly Gly Ala Asp Thr  Val Ala Arg
    2600                 2605                 2610

Ile Val  Leu Gly Ser Ser Pro  Ala Glu Ile Leu Val  Gly Asn Trp
    2615                 2620                 2625

Arg Thr  Pro Ser Lys Lys Val  Gly Ser Asp Thr Ile  Thr Leu His
    2630                 2635                 2640

Arg Lys  Ile Ser Ala Lys Ser  Asn Pro Phe Leu Glu  Asp His Val
    2645                 2650                 2655

Ile Gln  Gly Arg Arg Val Leu  Pro Met Thr Leu Ala  Ile Gly Ser
    2660                 2665                 2670

Leu Ala  Glu Thr Cys Leu Gly  Leu Phe Pro Gly Tyr  Ser Leu Trp
    2675                 2680                 2685

Ala Ile  Asp Asp Ala Gln Leu  Phe Lys Gly Val Thr  Val Asp Gly
    2690                 2695                 2700
```

```
Asp Val  Asn Cys Glu Val Thr  Leu Thr Pro Ser Thr  Ala Pro Ser
    2705                 2710                 2715

Gly Arg  Val Asn Val Gln Ala  Thr Leu Lys Thr Phe  Ser Ser Gly
    2720                 2725                 2730

Lys Leu  Val Pro Ala Tyr Arg  Ala Val Ile Val Leu  Ser Asn Gln
    2735                 2740                 2745

Gly Ala  Pro Pro Ala Asn Ala  Thr Met Gln Pro Pro  Ser Leu Asp
    2750                 2755                 2760

Ala Asp  Pro Ala Leu Gln Gly  Ser Val Tyr Asp Gly  Lys Thr Leu
    2765                 2770                 2775

Phe His  Gly Pro Ala Phe Arg  Gly Ile Asp Asp Val  Leu Ser Cys
    2780                 2785                 2790

Thr Lys  Ser Gln Leu Val Ala  Lys Cys Ser Ala Val  Pro Gly Ser
    2795                 2800                 2805

Asp Ala  Ala Arg Gly Glu Phe  Ala Thr Asp Thr Asp  Ala His Asp
    2810                 2815                 2820

Pro Phe  Val Asn Asp Leu Ala  Phe Gln Ala Met Leu  Val Trp Val
    2825                 2830                 2835

Arg Arg  Thr Leu Gly Gln Ala  Ala Leu Pro Asn Ser  Ile Gln Arg
    2840                 2845                 2850

Ile Val  Gln His Arg Pro Val  Pro Gln Asp Lys Pro  Phe Tyr Ile
    2855                 2860                 2865

Thr Leu  Arg Ser Asn Gln Ser  Gly Gly His Ser Gln  His Lys His
    2870                 2875                 2880

Ala Leu  Gln Phe His Asn Glu  Gln Gly Asp Leu Phe  Ile Asp Val
    2885                 2890                 2895

Gln Ala  Ser Val Ile Ala Thr  Asp Ser Leu Ala Phe
    2900                 2905                 2910


<210>  3
<211>  6180
<212>  DNA
<213>  Schizochytrium sp.


<220>
<221>  CDS
<222>  (1)..(6180)
```

<400> 3

```
atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag       48
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1               5                   10                  15

cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc       96
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
            20                  25                  30

aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag      144
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
        35                  40                  45

gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac      192
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
    50                  55                  60

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac      240
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65                  70                  75                  80

ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc      288
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
                85                  90                  95

aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca      336
Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
            100                 105                 110

cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc      384
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
        115                 120                 125

cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc      432
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
    130                 135                 140

ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag      480
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145                 150                 155                 160

tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc      528
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
                165                 170                 175

tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc      576
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
            180                 185                 190

gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat      624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
        195                 200                 205

cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc      672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
    210                 215                 220

ctg ccg gag ccc ttt ttc atc ctt tcg ggc ttt tcc acc ttc cag gcc      720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225                 230                 235                 240

atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac      768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
                245                 250                 255

agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag      816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
            260                 265                 270
```

```
cgt ctc gat gat gcc atc cgc gac ggc gac cac atc tac ggc acc ctt      864
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
        275                 280                 285

ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc      912
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
    290                 295                 300

ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att      960
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305                 310                 315                 320

aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc     1008
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
                325                 330                 335

acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt     1056
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
            340                 345                 350

gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac     1104
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
        355                 360                 365

acc ctc gtc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc     1152
Thr Leu Val Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
    370                 375                 380

atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc     1200
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385                 390                 395                 400

aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag     1248
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
                405                 410                 415

acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt     1296
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
            420                 425                 430

ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc     1344
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
        435                 440                 445

gcc tgc acg ggc cac gac tcc att tct gcg ctc tcg gcc cgc tgc ggc     1392
Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly
    450                 455                 460

ggt gaa agc aac atg cgc atc gcc atc act ggt atg gac gcc acc ttt     1440
Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe
465                 470                 475                 480

ggc gct ctc aag gga ctc gac gcc ttc gag cgc gcc att tac acc ggc     1488
Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly
                485                 490                 495

gct cac ggt gcc atc cca ctc cca gaa aag cgc tgg cgc ttt ctc ggc     1536
Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly
            500                 505                 510

aag gac aag gac ttt ctt gac ctc tgc ggc gtc aag gcc acc ccg cac     1584
Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His
        515                 520                 525

ggc tgc tac att gaa gat gtt gag gtc gac ttc cag cgc ctc cgc acg     1632
Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr
    530                 535                 540
```

```
ccc atg acc cct gaa gac atg ctc ctc cct cag cag ctt ctg gcc gtc     1680
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val
545                 550                 555                 560

acc acc att gac cgc gcc atc ctc gac tcg gga atg aaa aag ggt ggc     1728
Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly
                565                 570                 575

aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac ctc gag ctc tac cgt     1776
Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg
            580                 585                 590

cac cgt gct cgc gtc gct ctc aag gag cgc gtc cgc cct gaa gcc tcc     1824
His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser
        595                 600                 605

aag aag ctc aat gac atg atg cag tac att aac gac tgc ggc aca tcc     1872
Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser
    610                 615                 620

aca tcg tac acc tcg tac att ggc aac ctc gtc gcc acg cgc gtc tcg     1920
Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser
625                 630                 635                 640

tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg atc acc gag ggc aac     1968
Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn
                645                 650                 655

aac tcc gtc tac cgc tgc gcc gag ctc ggc aag tac ctc ctc gag acc     2016
Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr
            660                 665                 670

ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc gat ctc tgc ggc agt     2064
Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser
        675                 680                 685

gcc gaa aac ctt tac gtc aag tct cgc cgc ttc aag gtg tcc acc tcc     2112
Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser
    690                 695                 700

gat acc ccg cgc gcc agc ttt gac gcc gcc gcc gat ggc tac ttt gtc     2160
Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val
705                 710                 715                 720

ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt gag act agc tgc acc     2208
Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr
                725                 730                 735

aag gac gac cgt atc tac gct tgc atg gat gcc atc gtc cct ggc aac     2256
Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn
            740                 745                 750

gtc cct agc gcc tgc ttg cgc gag gcc ctc gac cag gcg cgc gtc aag     2304
Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys
        755                 760                 765

ccg ggc gat atc gag atg ctc gag ctc agc gcc gac tcc gcc cgc cac     2352
Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His
    770                 775                 780

ctc aag gac ccg tcc gtc ctg ccc aag gag ctc act gcc gag gag gaa     2400
Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu
785                 790                 795                 800

atc ggc ggc ctt cag acg atc ctt cgt gac gat gac aag ctc ccg cgc     2448
Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg
                805                 810                 815
```

```
aac gtc gca acg ggc agt gtc aag gcc acc gtc ggt gac acc ggt tat     2496
Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr
            820                 825                 830

gcc tct ggt gct gcc agc ctc atc aag gct gcg ctt tgc atc tac aac     2544
Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn
        835                 840                 845

cgc tac ctg ccc agc aac ggc gac gac tgg gat gaa ccc gcc cct gag     2592
Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu
    850                 855                 860

gcg ccc tgg gac agc acc ctc ttt gcg tgc cag acc tcg cgc gct tgg     2640
Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp
865                 870                 875                 880

ctc aag aac cct ggc gag cgt cgc tat gcg gcc gtc tcg ggc gtc tcc     2688
Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser
                885                 890                 895

gag acg cgc tcg tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac     2736
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His
            900                 905                 910

tac gag cgc gag aac cgc atc tcg ctc gac gag gag gcg ccc aag ctc     2784
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu
        915                 920                 925

att gtg ctt cgc gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac     2832
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp
    930                 935                 940

aag atc cgc gag cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag     2880
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu
945                 950                 955                 960

tcc gag ctc aag gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc     2928
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly
                965                 970                 975

gag acc ctt gcc cag gat gcc gct tct tca ggc tcg caa aag ccc ctc     2976
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu
            980                 985                 990

gct ctc agc ctc gtc tcc acg ccc  tcc aag ctc cag cgc  gag gtc gag   3024
Ala Leu Ser Leu Val Ser Thr Pro  Ser Lys Leu Gln Arg  Glu Val Glu
        995                 1000                1005

ctc gcg  gcc aag ggt atc ccg  cgc tgc ctc aag atg  cgc cgc gat      3069
Leu Ala  Ala Lys Gly Ile Pro  Arg Cys Leu Lys Met  Arg Arg Asp
    1010                1015                1020

tgg agc  tcc cct gct ggc agc  cgc tac gcg cct gag  ccg ctc gcc      3114
Trp Ser  Ser Pro Ala Gly Ser  Arg Tyr Ala Pro Glu  Pro Leu Ala
    1025                1030                1035

agc gac  cgc gtc gcc ttc atg  tac ggc gaa ggt cgc  agc cct tac      3159
Ser Asp  Arg Val Ala Phe Met  Tyr Gly Glu Gly Arg  Ser Pro Tyr
    1040                1045                1050

tac ggc  atc acc caa gac att  cac cgc att tgg ccc  gaa ctc cac      3204
Tyr Gly  Ile Thr Gln Asp Ile  His Arg Ile Trp Pro  Glu Leu His
    1055                1060                1065

gag gtc  atc aac gaa aag acg  aac cgt ctc tgg gcc  gaa ggc gac      3249
Glu Val  Ile Asn Glu Lys Thr  Asn Arg Leu Trp Ala  Glu Gly Asp
    1070                1075                1080
```

```
cgc tgg  gtc atg ccg cgc gcc  agc ttc aag tcg gag  ctc gag agc      3294
Arg Trp  Val Met Pro Arg Ala  Ser Phe Lys Ser Glu  Leu Glu Ser
    1085                 1090                 1095

cag cag  caa gag ttt gat cgc  aac atg att gaa atg  ttc cgt ctt      3339
Gln Gln  Gln Glu Phe Asp Arg  Asn Met Ile Glu Met  Phe Arg Leu
    1100                 1105                 1110

gga atc  ctc acc tca att gcc  ttc acc aat ctg gcg  cgc gac gtt      3384
Gly Ile  Leu Thr Ser Ile Ala  Phe Thr Asn Leu Ala  Arg Asp Val
    1115                 1120                 1125

ctc aac  atc acg ccc aag gcc  gcc ttt ggc ctc agt  ctt ggc gag      3429
Leu Asn  Ile Thr Pro Lys Ala  Ala Phe Gly Leu Ser  Leu Gly Glu
    1130                 1135                 1140

att tcc  atg att ttt gcc ttt  tcc aag aag aac ggt  ctc atc tcc      3474
Ile Ser  Met Ile Phe Ala Phe  Ser Lys Lys Asn Gly  Leu Ile Ser
    1145                 1150                 1155

gac cag  ctc acc aag gat ctt  cgc gag tcc gac gtg  tgg aac aag      3519
Asp Gln  Leu Thr Lys Asp Leu  Arg Glu Ser Asp Val  Trp Asn Lys
    1160                 1165                 1170

gct ctg  gcc gtt gaa ttt aat  gcg ctg cgc gag gcc  tgg ggc att      3564
Ala Leu  Ala Val Glu Phe Asn  Ala Leu Arg Glu Ala  Trp Gly Ile
    1175                 1180                 1185

cca cag  agt gtc ccc aag gac  gag ttc tgg caa ggc  tac att gtg      3609
Pro Gln  Ser Val Pro Lys Asp  Glu Phe Trp Gln Gly  Tyr Ile Val
    1190                 1195                 1200

cgc ggc  acc aag cag gat atc  gag gcg gcc atc gcc  ccg gac agc      3654
Arg Gly  Thr Lys Gln Asp Ile  Glu Ala Ala Ile Ala  Pro Asp Ser
    1205                 1210                 1215

aag tac  gtg cgc ctc acc atc  atc aat gat gcc aac  acc gcc ctc      3699
Lys Tyr  Val Arg Leu Thr Ile  Ile Asn Asp Ala Asn  Thr Ala Leu
    1220                 1225                 1230

att agc  ggc aag ccc gac gcc  tgc aag gct gcg atc  gcg cgt ctc      3744
Ile Ser  Gly Lys Pro Asp Ala  Cys Lys Ala Ala Ile  Ala Arg Leu
    1235                 1240                 1245

ggt ggc  aac att cct gcg ctt  ccc gtg acc cag ggc  atg tgc ggc      3789
Gly Gly  Asn Ile Pro Ala Leu  Pro Val Thr Gln Gly  Met Cys Gly
    1250                 1255                 1260

cac tgc  ccc gag gtg gga cct  tat acc aag gat atc  gcc aag atc      3834
His Cys  Pro Glu Val Gly Pro  Tyr Thr Lys Asp Ile  Ala Lys Ile
    1265                 1270                 1275

cat gcc  aac ctt gag ttc ccc  gtt gtc gac ggc ctt  gac ctc tgg      3879
His Ala  Asn Leu Glu Phe Pro  Val Val Asp Gly Leu  Asp Leu Trp
    1280                 1285                 1290

acc aca  atc aac cag aag cgc  ctc gtg cca cgc gcc  acg ggc gcc      3924
Thr Thr  Ile Asn Gln Lys Arg  Leu Val Pro Arg Ala  Thr Gly Ala
    1295                 1300                 1305

aag gac  gaa tgg gcc cct tct  tcc ttt ggc gag tac  gcc ggc cag      3969
Lys Asp  Glu Trp Ala Pro Ser  Ser Phe Gly Glu Tyr  Ala Gly Gln
    1310                 1315                 1320

ctc tac  gag aag cag gct aac  ttc ccc caa atc gtc  gag acc att      4014
Leu Tyr  Glu Lys Gln Ala Asn  Phe Pro Gln Ile Val  Glu Thr Ile
    1325                 1330                 1335
```

```
tac aag  caa aac tac gac gtc  ttt gtc gag gtt ggg  ccc aac aac      4059
Tyr Lys  Gln Asn Tyr Asp Val  Phe Val Glu Val Gly  Pro Asn Asn
    1340                 1345                 1350

cac cgt  agc acc gca gtg cgc  acc acg ctt ggt ccc  cag cgc aac      4104
His Arg  Ser Thr Ala Val Arg  Thr Thr Leu Gly Pro  Gln Arg Asn
    1355                 1360                 1365

cac ctt  gct ggc gcc atc gac  aag cag aac gag gat  gct tgg acg      4149
His Leu  Ala Gly Ala Ile Asp  Lys Gln Asn Glu Asp  Ala Trp Thr
    1370                 1375                 1380

acc atc  gtc aag ctt gtg gct  tcg ctc aag gcc cac  ctt gtt cct      4194
Thr Ile  Val Lys Leu Val Ala  Ser Leu Lys Ala His  Leu Val Pro
    1385                 1390                 1395

ggc gtc  acg atc tcg ccg ctg  tac cac tcc aag ctt  gtg gcg gag      4239
Gly Val  Thr Ile Ser Pro Leu  Tyr His Ser Lys Leu  Val Ala Glu
    1400                 1405                 1410

gct gag  gct tgc tac gct gcg  ctc tgc aag ggt gaa  aag ccc aag      4284
Ala Glu  Ala Cys Tyr Ala Ala  Leu Cys Lys Gly Glu  Lys Pro Lys
    1415                 1420                 1425

aag aac  aag ttt gtg cgc aag  att cag ctc aac ggt  cgc ttc aac      4329
Lys Asn  Lys Phe Val Arg Lys  Ile Gln Leu Asn Gly  Arg Phe Asn
    1430                 1435                 1440

agc aag  gcg gac ccc atc tcc  tcg gcc gat ctt gcc  agc ttt ccg      4374
Ser Lys  Ala Asp Pro Ile Ser  Ser Ala Asp Leu Ala  Ser Phe Pro
    1445                 1450                 1455

cct gcg  gac cct gcc att gaa  gcc gcc atc tcg agc  cgc atc atg      4419
Pro Ala  Asp Pro Ala Ile Glu  Ala Ala Ile Ser Ser  Arg Ile Met
    1460                 1465                 1470

aag cct  gtc gct ccc aag ttc  tac gcg cgt ctc aac  att gac gag      4464
Lys Pro  Val Ala Pro Lys Phe  Tyr Ala Arg Leu Asn  Ile Asp Glu
    1475                 1480                 1485

cag gac  gag acc cga gat ccg  atc ctc aac aag gac  aac gcg ccg      4509
Gln Asp  Glu Thr Arg Asp Pro  Ile Leu Asn Lys Asp  Asn Ala Pro
    1490                 1495                 1500

tct tct  tct tct tct tct tct  tct tct tct tct tct  tct tct tct      4554
Ser Ser  Ser Ser Ser Ser Ser  Ser Ser Ser Ser Ser  Ser Ser Ser
    1505                 1510                 1515

ccg tcg  cct gct cct tcg gcc  ccc gtg caa aag aag  gct gct ccc      4599
Pro Ser  Pro Ala Pro Ser Ala  Pro Val Gln Lys Lys  Ala Ala Pro
    1520                 1525                 1530

gcc gcg  gag acc aag gct gtt  gct tcg gct gac gca  ctt cgc agt      4644
Ala Ala  Glu Thr Lys Ala Val  Ala Ser Ala Asp Ala  Leu Arg Ser
    1535                 1540                 1545

gcc ctg  ctc gat ctc gac agt  atg ctt gcg ctg agc  tct gcc agt      4689
Ala Leu  Leu Asp Leu Asp Ser  Met Leu Ala Leu Ser  Ser Ala Ser
    1550                 1555                 1560

gcc tcc  ggc aac ctt gtt gag  act gcg cct agc gac  gcc tcg gtc      4734
Ala Ser  Gly Asn Leu Val Glu  Thr Ala Pro Ser Asp  Ala Ser Val
    1565                 1570                 1575

att gtg  ccg ccc tgc aac att  gcg gat ctc ggc agc  cgc gcc ttc      4779
Ile Val  Pro Pro Cys Asn Ile  Ala Asp Leu Gly Ser  Arg Ala Phe
    1580                 1585                 1590
```

```
atg aaa  acg tac ggt gtt tcg  gcg cct ctg tac acg  ggc gcc atg      4824
Met Lys  Thr Tyr Gly Val Ser  Ala Pro Leu Tyr Thr  Gly Ala Met
    1595                 1600                 1605

gcc aag  ggc att gcc tct gcg  gac ctc gtc att gcc  gcc ggc cgc      4869
Ala Lys  Gly Ile Ala Ser Ala  Asp Leu Val Ile Ala  Ala Gly Arg
    1610                 1615                 1620

cag ggc  atc ctt gcg tcc ttt  ggc gcc ggc gga ctt  ccc atg cag      4914
Gln Gly  Ile Leu Ala Ser Phe  Gly Ala Gly Gly Leu  Pro Met Gln
    1625                 1630                 1635

gtt gtg  cgt gag tcc atc gaa  aag att cag gcc gcc  ctg ccc aat      4959
Val Val  Arg Glu Ser Ile Glu  Lys Ile Gln Ala Ala  Leu Pro Asn
    1640                 1645                 1650

ggc ccg  tac gct gtc aac ctt  atc cat tct ccc ttt  gac agc aac      5004
Gly Pro  Tyr Ala Val Asn Leu  Ile His Ser Pro Phe  Asp Ser Asn
    1655                 1660                 1665

ctc gaa  aag ggc aat gtc gat  ctc ttc ctc gag aag  ggt gtc acc      5049
Leu Glu  Lys Gly Asn Val Asp  Leu Phe Leu Glu Lys  Gly Val Thr
    1670                 1675                 1680

ttt gtc  gag gcc tcg gcc ttt  atg acg ctc acc ccg  cag gtc gtg      5094
Phe Val  Glu Ala Ser Ala Phe  Met Thr Leu Thr Pro  Gln Val Val
    1685                 1690                 1695

cgg tac  cgc gcg gct ggc ctc  acg cgc aac gcc gac  ggc tcg gtc      5139
Arg Tyr  Arg Ala Ala Gly Leu  Thr Arg Asn Ala Asp  Gly Ser Val
    1700                 1705                 1710

aac atc  cgc aac cgt atc att  ggc aag gtc tcg cgc  acc gag ctc      5184
Asn Ile  Arg Asn Arg Ile Ile  Gly Lys Val Ser Arg  Thr Glu Leu
    1715                 1720                 1725

gcc gag  atg ttc atg cgt cct  gcg ccc gag cac ctt  ctt cag aag      5229
Ala Glu  Met Phe Met Arg Pro  Ala Pro Glu His Leu  Leu Gln Lys
    1730                 1735                 1740

ctc att  gct tcc ggc gag atc  aac cag gag cag gcc  gag ctc gcc      5274
Leu Ile  Ala Ser Gly Glu Ile  Asn Gln Glu Gln Ala  Glu Leu Ala
    1745                 1750                 1755

cgc cgt  gtt ccc gtc gct gac  gac atc gcg gtc gaa  gct gac tcg      5319
Arg Arg  Val Pro Val Ala Asp  Asp Ile Ala Val Glu  Ala Asp Ser
    1760                 1765                 1770

ggt ggc  cac acc gac aac cgc  ccc atc cac gtc att  ctg ccc ctc      5364
Gly Gly  His Thr Asp Asn Arg  Pro Ile His Val Ile  Leu Pro Leu
    1775                 1780                 1785

atc atc  aac ctt cgc gac cgc  ctt cac cgc gag tgc  ggc tac ccg      5409
Ile Ile  Asn Leu Arg Asp Arg  Leu His Arg Glu Cys  Gly Tyr Pro
    1790                 1795                 1800

gcc aac  ctt cgc gtc cgt gtg  ggc gcc ggc ggt ggc  att ggg tgc      5454
Ala Asn  Leu Arg Val Arg Val  Gly Ala Gly Gly Gly  Ile Gly Cys
    1805                 1810                 1815

ccc cag  gcg gcg ctg gcc acc  ttc aac atg ggt gcc  tcc ttt att      5499
Pro Gln  Ala Ala Leu Ala Thr  Phe Asn Met Gly Ala  Ser Phe Ile
    1820                 1825                 1830

gtc acc  ggc acc gtg aac cag  gtc gcc aag cag tcg  ggc acg tgc      5544
Val Thr  Gly Thr Val Asn Gln  Val Ala Lys Gln Ser  Gly Thr Cys
    1835                 1840                 1845
```

```
gac aat  gtg cgc aag cag ctc  gcg aag gcc act tac  tcg gac gta      5589
Asp Asn  Val Arg Lys Gln Leu  Ala Lys Ala Thr Tyr  Ser Asp Val
    1850                 1855                 1860

tgc atg  gcc ccg gct gcc gac  atg ttc gag gaa ggc  gtc aag ctt      5634
Cys Met  Ala Pro Ala Ala Asp  Met Phe Glu Glu Gly  Val Lys Leu
    1865                 1870                 1875

cag gtc  ctc aag aag gga acc  atg ttt ccc tcg cgc  gcc aac aag      5679
Gln Val  Leu Lys Lys Gly Thr  Met Phe Pro Ser Arg  Ala Asn Lys
    1880                 1885                 1890

ctc tac  gag ctc ttt tgc aag  tac gac tcg ttc gag  tcc atg ccc      5724
Leu Tyr  Glu Leu Phe Cys Lys  Tyr Asp Ser Phe Glu  Ser Met Pro
    1895                 1900                 1905

ccc gca  gag ctt gcg cgc gtc  gag aag cgc atc ttc  agc cgc gcg      5769
Pro Ala  Glu Leu Ala Arg Val  Glu Lys Arg Ile Phe  Ser Arg Ala
    1910                 1915                 1920

ctc gaa  gag gtc tgg gac gag  acc aaa aac ttt tac  att aac cgt      5814
Leu Glu  Glu Val Trp Asp Glu  Thr Lys Asn Phe Tyr  Ile Asn Arg
    1925                 1930                 1935

ctt cac  aac ccg gag aag atc  cag cgc gcc gag cgc  gac ccc aag      5859
Leu His  Asn Pro Glu Lys Ile  Gln Arg Ala Glu Arg  Asp Pro Lys
    1940                 1945                 1950

ctc aag  atg tcg ctg tgc ttt  cgc tgg tac ctg agc  ctg gcg agc      5904
Leu Lys  Met Ser Leu Cys Phe  Arg Trp Tyr Leu Ser  Leu Ala Ser
    1955                 1960                 1965

cgc tgg  gcc aac act gga gct  tcc gat cgc gtc atg  gac tac cag      5949
Arg Trp  Ala Asn Thr Gly Ala  Ser Asp Arg Val Met  Asp Tyr Gln
    1970                 1975                 1980

gtc tgg  tgc ggt cct gcc att  ggt tcc ttc aac gat  ttc atc aag      5994
Val Trp  Cys Gly Pro Ala Ile  Gly Ser Phe Asn Asp  Phe Ile Lys
    1985                 1990                 1995

gga act  tac ctt gat ccg gcc  gtc gca aac gag tac  ccg tgc gtc      6039
Gly Thr  Tyr Leu Asp Pro Ala  Val Ala Asn Glu Tyr  Pro Cys Val
    2000                 2005                 2010

gtt cag  att aac aag cag atc  ctt cgt gga gcg tgc  ttc ttg cgc      6084
Val Gln  Ile Asn Lys Gln Ile  Leu Arg Gly Ala Cys  Phe Leu Arg
    2015                 2020                 2025

cgt ctc  gaa att ctg cgc aac  gca cgc ctt tcc gat  ggc gct gcc      6129
Arg Leu  Glu Ile Leu Arg Asn  Ala Arg Leu Ser Asp  Gly Ala Ala
    2030                 2035                 2040

gct ctt  gtg gcc agc atc gat  gac aca tac gtc ccg  gcc gag aag      6174
Ala Leu  Val Ala Ser Ile Asp  Asp Thr Tyr Val Pro  Ala Glu Lys
    2045                 2050                 2055

ctg taa                                                          6180
Leu
```

<210> 4
<211> 2059
<212> PRT
<213> Schizochytrium sp.

<400> 4

```
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1               5                   10                  15

Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
            20                  25                  30

Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
        35                  40                  45

Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
    50                  55                  60

Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65                  70                  75                  80

Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
                85                  90                  95

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
            100                 105                 110

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
        115                 120                 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
    130                 135                 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145                 150                 155                 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
                165                 170                 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
            180                 185                 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
        195                 200                 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
    210                 215                 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225                 230                 235                 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
                245                 250                 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
            260                 265                 270
```

```
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
        275                 280                 285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
    290                 295                 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305                 310                 315                 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
                325                 330                 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
            340                 345                 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
        355                 360                 365

Thr Leu Val Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
    370                 375                 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385                 390                 395                 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
                405                 410                 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
            420                 425                 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
        435                 440                 445

Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly
    450                 455                 460

Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe
465                 470                 475                 480

Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly
                485                 490                 495

Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly
            500                 505                 510

Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His
        515                 520                 525

Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr
    530                 535                 540
```

```
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val
545                 550                 555                 560

Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly
                565                 570                 575

Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg
            580                 585                 590

His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser
        595                 600                 605

Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser
    610                 615                 620

Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser
625                 630                 635                 640

Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn
                645                 650                 655

Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr
            660                 665                 670

Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser
        675                 680                 685

Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser
    690                 695                 700

Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val
705                 710                 715                 720

Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr
                725                 730                 735

Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn
            740                 745                 750

Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys
        755                 760                 765

Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His
    770                 775                 780

Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu
785                 790                 795                 800

Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg
                805                 810                 815
```

```
Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr
            820                 825                 830

Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn
        835                 840                 845

Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu
    850                 855                 860

Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp
865                 870                 875                 880

Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser
                885                 890                 895

Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His
            900                 905                 910

Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu
        915                 920                 925

Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp
    930                 935                 940

Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu
945                 950                 955                 960

Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly
                965                 970                 975

Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu
            980                 985                 990

Ala Leu Ser Leu Val Ser Thr Pro  Ser Lys Leu Gln Arg  Glu Val Glu
        995                 1000                1005

Leu Ala  Ala Lys Gly Ile Pro  Arg Cys Leu Lys Met  Arg Arg Asp
    1010                1015                1020

Trp Ser  Ser Pro Ala Gly Ser  Arg Tyr Ala Pro Glu  Pro Leu Ala
    1025                1030                1035

Ser Asp  Arg Val Ala Phe Met  Tyr Gly Glu Gly Arg  Ser Pro Tyr
    1040                1045                1050

Tyr Gly  Ile Thr Gln Asp Ile  His Arg Ile Trp Pro  Glu Leu His
    1055                1060                1065

Glu Val  Ile Asn Glu Lys Thr  Asn Arg Leu Trp Ala  Glu Gly Asp
    1070                1075                1080
```

```
Arg Trp  Val Met Pro Arg Ala  Ser Phe Lys Ser Glu  Leu Glu Ser
    1085                 1090                 1095

Gln Gln  Gln Glu Phe Asp Arg  Asn Met Ile Glu Met  Phe Arg Leu
    1100                 1105                 1110

Gly Ile  Leu Thr Ser Ile Ala  Phe Thr Asn Leu Ala  Arg Asp Val
    1115                 1120                 1125

Leu Asn  Ile Thr Pro Lys Ala  Ala Phe Gly Leu Ser  Leu Gly Glu
    1130                 1135                 1140

Ile Ser  Met Ile Phe Ala Phe  Ser Lys Lys Asn Gly  Leu Ile Ser
    1145                 1150                 1155

Asp Gln  Leu Thr Lys Asp Leu  Arg Glu Ser Asp Val  Trp Asn Lys
    1160                 1165                 1170

Ala Leu  Ala Val Glu Phe Asn  Ala Leu Arg Glu Ala  Trp Gly Ile
    1175                 1180                 1185

Pro Gln  Ser Val Pro Lys Asp  Glu Phe Trp Gln Gly  Tyr Ile Val
    1190                 1195                 1200

Arg Gly  Thr Lys Gln Asp Ile  Glu Ala Ala Ile Ala  Pro Asp Ser
    1205                 1210                 1215

Lys Tyr  Val Arg Leu Thr Ile  Ile Asn Asp Ala Asn  Thr Ala Leu
    1220                 1225                 1230

Ile Ser  Gly Lys Pro Asp Ala  Cys Lys Ala Ala Ile  Ala Arg Leu
    1235                 1240                 1245

Gly Gly  Asn Ile Pro Ala Leu  Pro Val Thr Gln Gly  Met Cys Gly
    1250                 1255                 1260

His Cys  Pro Glu Val Gly Pro  Tyr Thr Lys Asp Ile  Ala Lys Ile
    1265                 1270                 1275

His Ala  Asn Leu Glu Phe Pro  Val Val Asp Gly Leu  Asp Leu Trp
    1280                 1285                 1290

Thr Thr  Ile Asn Gln Lys Arg  Leu Val Pro Arg Ala  Thr Gly Ala
    1295                 1300                 1305

Lys Asp  Glu Trp Ala Pro Ser  Ser Phe Gly Glu Tyr  Ala Gly Gln
    1310                 1315                 1320

Leu Tyr  Glu Lys Gln Ala Asn  Phe Pro Gln Ile Val  Glu Thr Ile
    1325                 1330                 1335
```

```
Tyr Lys  Gln Asn Tyr Asp Val  Phe Val Glu Val Gly  Pro Asn Asn
    1340                1345                1350

His Arg  Ser Thr Ala Val Arg  Thr Thr Leu Gly Pro  Gln Arg Asn
    1355                1360                1365

His Leu  Ala Gly Ala Ile Asp  Lys Gln Asn Glu Asp  Ala Trp Thr
    1370                1375                1380

Thr Ile  Val Lys Leu Val Ala  Ser Leu Lys Ala His  Leu Val Pro
    1385                1390                1395

Gly Val  Thr Ile Ser Pro Leu  Tyr His Ser Lys Leu  Val Ala Glu
    1400                1405                1410

Ala Glu  Ala Cys Tyr Ala Ala  Leu Cys Lys Gly Glu  Lys Pro Lys
    1415                1420                1425

Lys Asn  Lys Phe Val Arg Lys  Ile Gln Leu Asn Gly  Arg Phe Asn
    1430                1435                1440

Ser Lys  Ala Asp Pro Ile Ser  Ser Ala Asp Leu Ala  Ser Phe Pro
    1445                1450                1455

Pro Ala  Asp Pro Ala Ile Glu  Ala Ala Ile Ser Ser  Arg Ile Met
    1460                1465                1470

Lys Pro  Val Ala Pro Lys Phe  Tyr Ala Arg Leu Asn  Ile Asp Glu
    1475                1480                1485

Gln Asp  Glu Thr Arg Asp Pro  Ile Leu Asn Lys Asp  Asn Ala Pro
    1490                1495                1500

Ser Ser  Ser Ser Ser Ser Ser  Ser Ser Ser Ser Ser  Ser Ser Ser
    1505                1510                1515

Pro Ser  Pro Ala Pro Ser Ala  Pro Val Gln Lys Lys  Ala Ala Pro
    1520                1525                1530

Ala Ala  Glu Thr Lys Ala Val  Ala Ser Ala Asp Ala  Leu Arg Ser
    1535                1540                1545

Ala Leu  Leu Asp Leu Asp Ser  Met Leu Ala Leu Ser  Ser Ala Ser
    1550                1555                1560

Ala Ser  Gly Asn Leu Val Glu  Thr Ala Pro Ser Asp  Ala Ser Val
    1565                1570                1575

Ile Val  Pro Pro Cys Asn Ile  Ala Asp Leu Gly Ser  Arg Ala Phe
    1580                1585                1590
```

```
Met Lys  Thr Tyr Gly Val Ser  Ala Pro Leu Tyr Thr  Gly Ala Met
    1595                 1600                 1605

Ala Lys  Gly Ile Ala Ser Ala  Asp Leu Val Ile Ala  Ala Gly Arg
    1610                 1615                 1620

Gln Gly  Ile Leu Ala Ser Phe  Gly Ala Gly Gly Leu  Pro Met Gln
    1625                 1630                 1635

Val Val  Arg Glu Ser Ile Glu  Lys Ile Gln Ala Ala  Leu Pro Asn
    1640                 1645                 1650

Gly Pro  Tyr Ala Val Asn Leu  Ile His Ser Pro Phe  Asp Ser Asn
    1655                 1660                 1665

Leu Glu  Lys Gly Asn Val Asp  Leu Phe Leu Glu Lys  Gly Val Thr
    1670                 1675                 1680

Phe Val  Glu Ala Ser Ala Phe  Met Thr Leu Thr Pro  Gln Val Val
    1685                 1690                 1695

Arg Tyr  Arg Ala Ala Gly Leu  Thr Arg Asn Ala Asp  Gly Ser Val
    1700                 1705                 1710

Asn Ile  Arg Asn Arg Ile Ile  Gly Lys Val Ser Arg  Thr Glu Leu
    1715                 1720                 1725

Ala Glu  Met Phe Met Arg Pro  Ala Pro Glu His Leu  Leu Gln Lys
    1730                 1735                 1740

Leu Ile  Ala Ser Gly Glu Ile  Asn Gln Glu Gln Ala  Glu Leu Ala
    1745                 1750                 1755

Arg Arg  Val Pro Val Ala Asp  Asp Ile Ala Val Glu  Ala Asp Ser
    1760                 1765                 1770

Gly Gly  His Thr Asp Asn Arg  Pro Ile His Val Ile  Leu Pro Leu
    1775                 1780                 1785

Ile Ile  Asn Leu Arg Asp Arg  Leu His Arg Glu Cys  Gly Tyr Pro
    1790                 1795                 1800

Ala Asn  Leu Arg Val Arg Val  Gly Ala Gly Gly Gly  Ile Gly Cys
    1805                 1810                 1815

Pro Gln  Ala Ala Leu Ala Thr  Phe Asn Met Gly Ala  Ser Phe Ile
    1820                 1825                 1830

Val Thr  Gly Thr Val Asn Gln  Val Ala Lys Gln Ser  Gly Thr Cys
    1835                 1840                 1845
```

```
Asp Asn  Val Arg Lys Gln Leu  Ala Lys Ala Thr Tyr  Ser Asp Val
    1850                 1855                 1860

Cys Met  Ala Pro Ala Ala Asp  Met Phe Glu Glu Gly  Val Lys Leu
    1865                 1870                 1875

Gln Val  Leu Lys Lys Gly Thr  Met Phe Pro Ser Arg  Ala Asn Lys
    1880                 1885                 1890

Leu Tyr  Glu Leu Phe Cys Lys  Tyr Asp Ser Phe Glu  Ser Met Pro
    1895                 1900                 1905

Pro Ala  Glu Leu Ala Arg Val  Glu Lys Arg Ile Phe  Ser Arg Ala
    1910                 1915                 1920

Leu Glu  Glu Val Trp Asp Glu  Thr Lys Asn Phe Tyr  Ile Asn Arg
    1925                 1930                 1935

Leu His  Asn Pro Glu Lys Ile  Gln Arg Ala Glu Arg  Asp Pro Lys
    1940                 1945                 1950

Leu Lys  Met Ser Leu Cys Phe  Arg Trp Tyr Leu Ser  Leu Ala Ser
    1955                 1960                 1965

Arg Trp  Ala Asn Thr Gly Ala  Ser Asp Arg Val Met  Asp Tyr Gln
    1970                 1975                 1980

Val Trp  Cys Gly Pro Ala Ile  Gly Ser Phe Asn Asp  Phe Ile Lys
    1985                 1990                 1995

Gly Thr  Tyr Leu Asp Pro Ala  Val Ala Asn Glu Tyr  Pro Cys Val
    2000                 2005                 2010

Val Gln  Ile Asn Lys Gln Ile  Leu Arg Gly Ala Cys  Phe Leu Arg
    2015                 2020                 2025

Arg Leu  Glu Ile Leu Arg Asn  Ala Arg Leu Ser Asp  Gly Ala Ala
    2030                 2035                 2040

Ala Leu  Val Ala Ser Ile Asp  Asp Thr Tyr Val Pro  Ala Glu Lys
    2045                 2050                 2055

Leu
```

<210> 5
<211> 4509
<212> DNA
<213> Schizochytrium sp.

<220>

<221> CDS
<222> (1)..(4509)

<400> 5

```
atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc       48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1               5                  10                  15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa       96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
            20                  25                  30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag      144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
        35                  40                  45

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc      192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac      240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc      288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc      336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac      384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc      432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att      480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc      528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg      576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc      624
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc      672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag      720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg      768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
                245                 250                 255
```

```
gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa      816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
            260                 265                 270

ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac      864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
        275                 280                 285

cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc      912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
    290                 295                 300

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag      960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305                 310                 315                 320

gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag     1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
                325                 330                 335

atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac     1056
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
            340                 345                 350

ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa     1104
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag     1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
    370                 375                 380

atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac     1200
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385                 390                 395                 400

atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc     1248
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
                405                 410                 415

atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac     1296
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
            420                 425                 430

ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac     1344
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
        435                 440                 445

ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc     1392
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
    450                 455                 460

ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc     1440
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
465                 470                 475                 480

gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc     1488
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
                485                 490                 495

gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag     1536
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
            500                 505                 510

gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc     1584
Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
        515                 520                 525
```

```
tct ttt gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc     1632
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
    530                 535                 540

ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg     1680
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
545                 550                 555                 560

ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg     1728
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
                565                 570                 575

tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc     1776
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
            580                 585                 590

agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct     1824
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
        595                 600                 605

gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag     1872
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
    610                 615                 620

ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac     1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
625                 630                 635                 640

aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag     1968
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
                645                 650                 655

atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc     2016
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
            660                 665                 670

ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac     2064
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
        675                 680                 685

gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc     2112
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
    690                 695                 700

aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc     2160
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
705                 710                 715                 720

cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag     2208
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
                725                 730                 735

ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag     2256
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
            740                 745                 750

gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac     2304
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
        755                 760                 765

aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc     2352
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
    770                 775                 780

agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc     2400
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785                 790                 795                 800
```

```
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc     2448
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
                805                 810                 815

tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg     2496
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
            820                 825                 830

gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg     2544
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
        835                 840                 845

gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag     2592
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
    850                 855                 860

gcc atc gcc gcc cac gag gat ctc gct ggc aag cac ggc att gcc aac     2640
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn
865                 870                 875                 880

ccc acc ttt gtg cac gcc ccg ggc aag atc agc tgg aag tac cgc ggc     2688
Pro Thr Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly
                885                 890                 895

cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg     2736
Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val
            900                 905                 910

tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc     2784
Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe
        915                 920                 925

ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg     2832
Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val
    930                 935                 940

cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc tct     2880
Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala Ser
945                 950                 955                 960

gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc gct     2928
Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro Ala
                965                 970                 975

gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag acc     2976
Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys Thr
            980                 985                 990

gag ctc ctc gag ctc gat gcc ccg  ctc tac ctc tcg cag  gac ccg acc   3024
Glu Leu Leu Glu Leu Asp Ala Pro  Leu Tyr Leu Ser Gln  Asp Pro Thr
        995                 1000                1005

agc ggc  cag ctc aag aag cac  acc gac gtg gcc tcc  ggc cag gcc      3069
Ser Gly  Gln Leu Lys Lys His  Thr Asp Val Ala Ser  Gly Gln Ala
    1010                1015                1020

acc atc  gtg cag ccc tgc acg  ctc ggc gac ctc ggt  gac cgc tcc      3114
Thr Ile  Val Gln Pro Cys Thr  Leu Gly Asp Leu Gly  Asp Arg Ser
    1025                1030                1035

ttc atg  gag acc tac ggc gtc  gtc gcc ccg ctg tac  acg ggc gcc      3159
Phe Met  Glu Thr Tyr Gly Val  Val Ala Pro Leu Tyr  Thr Gly Ala
    1040                1045                1050

atg gcc  aag ggc att gcc tcg  gcg gac ctc gtc atc  gcc gcc ggc      3204
Met Ala  Lys Gly Ile Ala Ser  Ala Asp Leu Val Ile  Ala Ala Gly
    1055                1060                1065
```

```
aag cgc  aag atc ctc ggc tcc  ttt ggc gcc ggc ggc  ctc ccc atg      3249
Lys Arg  Lys Ile Leu Gly Ser  Phe Gly Ala Gly Gly  Leu Pro Met
    1070                 1075                 1080

cac cac  gtg cgc gcc gcc ctc  gag aag atc cag gcc  gcc ctg cct      3294
His His  Val Arg Ala Ala Leu  Glu Lys Ile Gln Ala  Ala Leu Pro
    1085                 1090                 1095

cag ggc  ccc tac gcc gtc aac  ctc atc cac tcg cct  ttt gac agc      3339
Gln Gly  Pro Tyr Ala Val Asn  Leu Ile His Ser Pro  Phe Asp Ser
    1100                 1105                 1110

aac ctc  gag aag ggc aac gtc  gat ctc ttc ctc gag  aag ggc gtc      3384
Asn Leu  Glu Lys Gly Asn Val  Asp Leu Phe Leu Glu  Lys Gly Val
    1115                 1120                 1125

act gtg  gtg gag gcc tcg gca  ttc atg acc ctc acc  ccg cag gtc      3429
Thr Val  Val Glu Ala Ser Ala  Phe Met Thr Leu Thr  Pro Gln Val
    1130                 1135                 1140

gtg cgc  tac cgc gcc gcc ggc  ctc tcg cgc aac gcc  gac ggt tcg      3474
Val Arg  Tyr Arg Ala Ala Gly  Leu Ser Arg Asn Ala  Asp Gly Ser
    1145                 1150                 1155

gtc aac  atc cgc aac cgc atc  atc ggc aag gtc tcg  cgc acc gag      3519
Val Asn  Ile Arg Asn Arg Ile  Ile Gly Lys Val Ser  Arg Thr Glu
    1160                 1165                 1170

ctc gcc  gag atg ttc atc cgc  ccg gcc ccg gag cac  ctc ctc gag      3564
Leu Ala  Glu Met Phe Ile Arg  Pro Ala Pro Glu His  Leu Leu Glu
    1175                 1180                 1185

aag ctc  atc gcc tcg ggc gag  atc acc cag gag cag  gcc gag ctc      3609
Lys Leu  Ile Ala Ser Gly Glu  Ile Thr Gln Glu Gln  Ala Glu Leu
    1190                 1195                 1200

gcg cgc  cgc gtt ccc gtc gcc  gac gat atc gct gtc  gag gct gac      3654
Ala Arg  Arg Val Pro Val Ala  Asp Asp Ile Ala Val  Glu Ala Asp
    1205                 1210                 1215

tcg ggc  ggc cac acc gac aac  cgc ccc atc cac gtc  atc ctc ccg      3699
Ser Gly  Gly His Thr Asp Asn  Arg Pro Ile His Val  Ile Leu Pro
    1220                 1225                 1230

ctc atc  atc aac ctc cgc aac  cgc ctg cac cgc gag  tgc ggc tac      3744
Leu Ile  Ile Asn Leu Arg Asn  Arg Leu His Arg Glu  Cys Gly Tyr
    1235                 1240                 1245

ccc gcg  cac ctc cgc gtc cgc  gtt ggc gcc ggc ggt  ggc gtc ggc      3789
Pro Ala  His Leu Arg Val Arg  Val Gly Ala Gly Gly  Gly Val Gly
    1250                 1255                 1260

tgc ccg  cag gcc gcc gcc gcc  gcg ctc acc atg ggc  gcc gcc ttc      3834
Cys Pro  Gln Ala Ala Ala Ala  Ala Leu Thr Met Gly  Ala Ala Phe
    1265                 1270                 1275

atc gtc  acc ggc act gtc aac  cag gtc gcc aag cag  tcc ggc acc      3879
Ile Val  Thr Gly Thr Val Asn  Gln Val Ala Lys Gln  Ser Gly Thr
    1280                 1285                 1290

tgc gac  aac gtg cgc aag cag  ctc tcg cag gcc acc  tac tcg gat      3924
Cys Asp  Asn Val Arg Lys Gln  Leu Ser Gln Ala Thr  Tyr Ser Asp
    1295                 1300                 1305

atc tgc  atg gcc ccg gcc gcc  gac atg ttc gag gag  ggc gtc aag      3969
Ile Cys  Met Ala Pro Ala Ala  Asp Met Phe Glu Glu  Gly Val Lys
    1310                 1315                 1320
```

```
ctc cag  gtc ctc aag aag gga  acc atg ttc ccc tcg  cgc gcc aac      4014
Leu Gln  Val Leu Lys Lys Gly  Thr Met Phe Pro Ser  Arg Ala Asn
    1325                 1330                 1335

aag ctc  tac gag ctc ttt tgc  aag tac gac tcc ttc  gac tcc atg      4059
Lys Leu  Tyr Glu Leu Phe Cys  Lys Tyr Asp Ser Phe  Asp Ser Met
    1340                 1345                 1350

cct cct  gcc gag ctc gag cgc  atc gag aag cgt atc  ttc aag cgc      4104
Pro Pro  Ala Glu Leu Glu Arg  Ile Glu Lys Arg Ile  Phe Lys Arg
    1355                 1360                 1365

gca ctc  cag gag gtc tgg gag  gag acc aag gac ttt  tac att aac      4149
Ala Leu  Gln Glu Val Trp Glu  Glu Thr Lys Asp Phe  Tyr Ile Asn
    1370                 1375                 1380

ggt ctc  aag aac ccg gag aag  atc cag cgc gcc gag  cac gac ccc      4194
Gly Leu  Lys Asn Pro Glu Lys  Ile Gln Arg Ala Glu  His Asp Pro
    1385                 1390                 1395

aag ctc  aag atg tcg ctc tgc  ttc cgc tgg tac ctt  ggt ctt gcc      4239
Lys Leu  Lys Met Ser Leu Cys  Phe Arg Trp Tyr Leu  Gly Leu Ala
    1400                 1405                 1410

agc cgc  tgg gcc aac atg ggc  gcc ccg gac cgc gtc  atg gac tac      4284
Ser Arg  Trp Ala Asn Met Gly  Ala Pro Asp Arg Val  Met Asp Tyr
    1415                 1420                 1425

cag gtc  tgg tgt ggc ccg gcc  att ggc gcc ttc aac  gac ttc atc      4329
Gln Val  Trp Cys Gly Pro Ala  Ile Gly Ala Phe Asn  Asp Phe Ile
    1430                 1435                 1440

aag ggc  acc tac ctc gac ccc  gct gtc tcc aac gag  tac ccc tgt      4374
Lys Gly  Thr Tyr Leu Asp Pro  Ala Val Ser Asn Glu  Tyr Pro Cys
    1445                 1450                 1455

gtc gtc  cag atc aac ctg caa  atc ctc cgt ggt gcc  tgc tac ctg      4419
Val Val  Gln Ile Asn Leu Gln  Ile Leu Arg Gly Ala  Cys Tyr Leu
    1460                 1465                 1470

cgc cgt  ctc aac gcc ctg cgc  aac gac ccg cgc att  gac ctc gag      4464
Arg Arg  Leu Asn Ala Leu Arg  Asn Asp Pro Arg Ile  Asp Leu Glu
    1475                 1480                 1485

acc gag  gat gct gcc ttt gtc  tac gag ccc acc aac  gcg ctc taa      4509
Thr Glu  Asp Ala Ala Phe Val  Tyr Glu Pro Thr Asn  Ala Leu
    1490                 1495                 1500
```

<210> 6
<211> 1502
<212> PRT
<213> Schizochytrium sp.

<400> 6

```
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1               5                  10                  15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
            20                  25                  30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
        35                  40                  45
```

```
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
                245                 250                 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
            260                 265                 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
        275                 280                 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
    290                 295                 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305                 310                 315                 320
```

```
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
                325                 330                 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
            340                 345                 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
    370                 375                 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385                 390                 395                 400

Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
                405                 410                 415

Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
            420                 425                 430

Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
        435                 440                 445

Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
    450                 455                 460

Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
465                 470                 475                 480

Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
                485                 490                 495

Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
            500                 505                 510

Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
        515                 520                 525

Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
    530                 535                 540

Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
545                 550                 555                 560

Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
                565                 570                 575

Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
            580                 585                 590
```

```
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
        595                 600                 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
    610                 615                 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
625                 630                 635                 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
                645                 650                 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
            660                 665                 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
        675                 680                 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
    690                 695                 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
705                 710                 715                 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
                725                 730                 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
            740                 745                 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
        755                 760                 765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
    770                 775                 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785                 790                 795                 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
                805                 810                 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
            820                 825                 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
        835                 840                 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
    850                 855                 860
```

```
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn
865                 870                 875                 880

Pro Thr Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly
                885                 890                 895

Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val
            900                 905                 910

Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe
        915                 920                 925

Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val
    930                 935                 940

Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala Ser
945                 950                 955                 960

Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro Ala
                965                 970                 975

Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys Thr
            980                 985                 990

Glu Leu Leu Glu Leu Asp Ala Pro  Leu Tyr Leu Ser Gln  Asp Pro Thr
        995                 1000                1005

Ser Gly  Gln Leu Lys Lys His  Thr Asp Val Ala Ser  Gly Gln Ala
    1010                1015                1020

Thr Ile  Val Gln Pro Cys Thr  Leu Gly Asp Leu Gly  Asp Arg Ser
    1025                1030                1035

Phe Met  Glu Thr Tyr Gly Val  Val Ala Pro Leu Tyr  Thr Gly Ala
    1040                1045                1050

Met Ala  Lys Gly Ile Ala Ser  Ala Asp Leu Val Ile  Ala Ala Gly
    1055                1060                1065

Lys Arg  Lys Ile Leu Gly Ser  Phe Gly Ala Gly Gly  Leu Pro Met
    1070                1075                1080

His His  Val Arg Ala Ala Leu  Glu Lys Ile Gln Ala  Ala Leu Pro
    1085                1090                1095

Gln Gly  Pro Tyr Ala Val Asn  Leu Ile His Ser Pro  Phe Asp Ser
    1100                1105                1110

Asn Leu  Glu Lys Gly Asn Val  Asp Leu Phe Leu Glu  Lys Gly Val
    1115                1120                1125
```

```
Thr Val  Val Glu Ala Ser Ala  Phe Met Thr Leu Thr  Pro Gln Val
    1130                1135                1140

Val Arg  Tyr Arg Ala Ala Gly  Leu Ser Arg Asn Ala  Asp Gly Ser
    1145                1150                1155

Val Asn  Ile Arg Asn Arg Ile  Ile Gly Lys Val Ser  Arg Thr Glu
    1160                1165                1170

Leu Ala  Glu Met Phe Ile Arg  Pro Ala Pro Glu His  Leu Leu Glu
    1175                1180                1185

Lys Leu  Ile Ala Ser Gly Glu  Ile Thr Gln Glu Gln  Ala Glu Leu
    1190                1195                1200

Ala Arg  Arg Val Pro Val Ala  Asp Asp Ile Ala Val  Glu Ala Asp
    1205                1210                1215

Ser Gly  Gly His Thr Asp Asn  Arg Pro Ile His Val  Ile Leu Pro
    1220                1225                1230

Leu Ile  Ile Asn Leu Arg Asn  Arg Leu His Arg Glu  Cys Gly Tyr
    1235                1240                1245

Pro Ala  His Leu Arg Val Arg  Val Gly Ala Gly Gly  Gly Val Gly
    1250                1255                1260

Cys Pro  Gln Ala Ala Ala Ala  Ala Leu Thr Met Gly  Ala Ala Phe
    1265                1270                1275

Ile Val  Thr Gly Thr Val Asn  Gln Val Ala Lys Gln  Ser Gly Thr
    1280                1285                1290

Cys Asp  Asn Val Arg Lys Gln  Leu Ser Gln Ala Thr  Tyr Ser Asp
    1295                1300                1305

Ile Cys  Met Ala Pro Ala Ala  Asp Met Phe Glu Glu  Gly Val Lys
    1310                1315                1320

Leu Gln  Val Leu Lys Lys Gly  Thr Met Phe Pro Ser  Arg Ala Asn
    1325                1330                1335

Lys Leu  Tyr Glu Leu Phe Cys  Lys Tyr Asp Ser Phe  Asp Ser Met
    1340                1345                1350

Pro Pro  Ala Glu Leu Glu Arg  Ile Glu Lys Arg Ile  Phe Lys Arg
    1355                1360                1365

Ala Leu  Gln Glu Val Trp Glu  Glu Thr Lys Asp Phe  Tyr Ile Asn
    1370                1375                1380
```

```
Gly Leu  Lys Asn Pro Glu Lys  Ile Gln Arg Ala Glu  His Asp Pro
    1385                 1390                 1395

Lys Leu  Lys Met Ser Leu Cys  Phe Arg Trp Tyr Leu  Gly Leu Ala
    1400                 1405                 1410

Ser Arg  Trp Ala Asn Met Gly  Ala Pro Asp Arg Val  Met Asp Tyr
    1415                 1420                 1425

Gln Val  Trp Cys Gly Pro Ala  Ile Gly Ala Phe Asn  Asp Phe Ile
    1430                 1435                 1440

Lys Gly  Thr Tyr Leu Asp Pro  Ala Val Ser Asn Glu  Tyr Pro Cys
    1445                 1450                 1455

Val Val  Gln Ile Asn Leu Gln  Ile Leu Arg Gly Ala  Cys Tyr Leu
    1460                 1465                 1470

Arg Arg  Leu Asn Ala Leu Arg  Asn Asp Pro Arg Ile  Asp Leu Glu
    1475                 1480                 1485

Thr Glu  Asp Ala Ala Phe Val  Tyr Glu Pro Thr Asn  Ala Leu
    1490                 1495                 1500


<210>  7
<211>  1500
<212>  DNA
<213>  Schizochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<400>  7
atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc       48
Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
1               5                  10                  15

att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg       96
Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
            20                  25                  30

cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat      144
Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
        35                  40                  45

ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag      192
Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
    50                  55                  60

acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag      240
Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65                  70                  75                  80

tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag      288
Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
                85                  90                  95

gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc      336
```

```
Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
            100                 105                 110

ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc      384
Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
        115                 120                 125

ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc      432
Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
    130                 135                 140

tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg      480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145                 150                 155                 160

ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc      528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
                165                 170                 175

aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac      576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
            180                 185                 190

gtc acc gcc ggt cgc tgc acc aac acc ttc aac ctc gac ggc atg aac      624
Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
        195                 200                 205

tgc gtt gtc gac gcc gca tgc gcc tcg tcc ctc atc gcc gtc aag gtc      672
Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
    210                 215                 220

gcc atc gac gag ctg ctc tac ggt gac tgc gac atg atg gtc acc ggt      720
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
225                 230                 235                 240

gcc acc tgc acg gat aac tcc atc ggc atg tac atg gcc ttc tcc aag      768
Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
                245                 250                 255

acc ccc gtg ttc tcc acg gac ccc agc gtg cgc gcc tac gac gaa aag      816
Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
            260                 265                 270

aca aag ggc atg ctc atc ggc gag ggc tcc gcc atg ctc gtc ctc aag      864
Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
        275                 280                 285

cgc tac gcc gac gcc gtc cgc gac ggc gat gag atc cac gct gtt att      912
Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
    290                 295                 300

cgc ggc tgc gcc tcc tcc agt gat ggc aag gcc gcc ggc atc tac acg      960
Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
305                 310                 315                 320

ccc acc att tcg ggc cag gag gag gcc ctc cgc cgc gcc tac aac cgc     1008
Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
                325                 330                 335

gcc tgt gtc gac ccg gcc acc gtc act ctc gtc gag ggt cac ggc acc     1056
Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
            340                 345                 350

ggt act ccc gtt ggc gac cgc atc gag ctc acc gcc ttg cgc aac ctc     1104
Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
        355                 360                 365

ttt gac aag gcc tac ggc gag ggc aac acc gaa aag gtc gct gtg ggc     1152
```

```
Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
    370                 375                 380

agc atc aag tcc agc atc ggc cat ctc aag gcc gtc gcc ggt ctc gcc     1200
Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385                 390                 395                 400

ggt atg atc aag gtc atc atg gcg ctc aag cac aag act ctc ccg ggc     1248
Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
                405                 410                 415

acc atc aac gtc gac aac cca ccc aac ctc tac gac aac acg ccc atc     1296
Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
            420                 425                 430

aac gag tcc tcg ctc tac att aac acc atg aac cgc ccc tgg ttc ccg     1344
Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
        435                 440                 445

ccc cct ggt gtg ccc cgc cgc gcc ggc att tcg agc ttt ggc ttt ggt     1392
Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
    450                 455                 460

ggc gcc aac tac cac gcc gtc ctc gag gag gcc gag ccc gag cac acg     1440
Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465                 470                 475                 480

acc gcg tac cgc ctc aac aag cgc ccg cag ccc gtg ctc atg atg gcc     1488
Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
                485                 490                 495

gcc acg ccc gcg                                                    1500
Ala Thr Pro Ala
            500
```

<210> 8
<211> 500
<212> PRT
<213> Schizochytrium sp.

<400> 8

```
Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
1               5                   10                  15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
            20                  25                  30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
        35                  40                  45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
    50                  55                  60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65                  70                  75                  80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
                85                  90                  95

Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
```

100 105 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
115 120 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
130 135 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145 150 155 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
165 170 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
180 185 190

Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
195 200 205

Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
210 215 220

Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
225 230 235 240

Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
245 250 255

Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
260 265 270

Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
275 280 285

Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
290 295 300

Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
305 310 315 320

Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
325 330 335

Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
340 345 350

Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
355 360 365

Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly

370 375 380

Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385 390 395 400

Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
405 410 415

Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
420 425 430

Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
435 440 445

Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
450 455 460

Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465 470 475 480

Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
485 490 495

Ala Thr Pro Ala
500

<210> 9
<211> 1278
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1278)

<400> 9
gat gtc acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc 48
Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1 5 10 15

cgc gcc aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc 96
Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser
20 25 30

ggc cag ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac 144
Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
35 40 45

tgg ccc cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc 192
Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
50 55 60

aag gtc gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc 240
Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
65 70 75 80

tac ccg cgc aag ccg tac gag cgt gag ccc gag cag gac cac aag aag 288
Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His Lys Lys

```
                85                  90                  95

atc tcc ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc      336
Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
            100                 105                 110

ggt gcc ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc      384
Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
        115                 120                 125

gcc ggc cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc      432
Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
    130                 135                 140

gtc gac cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc      480
Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
145                 150                 155                 160

atg ggc ggc aag gac gca ccg gcc acc ccc aag ggc tgc atg gcc gcc      528
Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
                165                 170                 175

gtc att ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc      576
Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
            180                 185                 190

tgg ctc ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc      624
Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser
        195                 200                 205

gtc gaa ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc      672
Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe
    210                 215                 220

cgc gtc gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg      720
Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met
225                 230                 235                 240

gag aac gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc      768
Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe
                245                 250                 255

cgc acc ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag      816
Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu
            260                 265                 270

acc tac ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc      864
Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser
        275                 280                 285

agc gtc aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg      912
Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala
    290                 295                 300

cgc atc ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc      960
Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val
305                 310                 315                 320

tcc gag acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac     1008
Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn
                325                 330                 335

ccg gcc tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc     1056
Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val
            340                 345                 350

cag ctc gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac     1104
Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp
```

```
        355                 360                 365

gcc ccc gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc     1152
Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
    370                 375                 380

ctc cgc ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc     1200
Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
385                 390                 395                 400

cgc gac gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc     1248
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
                405                 410                 415

gcc gca ccg ctc atc aag gcc ccg gag ccc                             1278
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
            420                 425


<210>  10
<211>  426
<212>  PRT
<213>  Schizochytrium sp.

<400>  10

Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1               5                   10                  15


Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser
            20                  25                  30


Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
        35                  40                  45


Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
    50                  55                  60


Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
65                  70                  75                  80


Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His Lys Lys
                85                  90                  95


Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
            100                 105                 110


Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
        115                 120                 125


Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
    130                 135                 140


Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
145                 150                 155                 160


Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
                165                 170                 175
```

```
Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
            180                 185                 190

Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser
        195                 200                 205

Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe
    210                 215                 220

Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met
225                 230                 235                 240

Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe
                245                 250                 255

Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu
            260                 265                 270

Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser
        275                 280                 285

Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala
    290                 295                 300

Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val
305                 310                 315                 320

Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn
                325                 330                 335

Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val
            340                 345                 350

Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp
        355                 360                 365

Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
    370                 375                 380

Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
385                 390                 395                 400

Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
                405                 410                 415

Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
            420                 425
```

<210> 11
<211> 5

<212> PRT
<213> Schizochytrium sp.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = any amino acid

<400> 11

Gly His Ser Xaa Gly
1 5

<210> 12
<211> 258
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(258)

<400> 12
gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc atg gag 48
Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1 5 10 15

gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc gag gct gac 96
Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
20 25 30

atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag 144
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
35 40 45

atc ctc tcc gag gtc cag gcc atg ctc aat gtc gag gcc aag gat gtc 192
Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
50 55 60

gat gcc ctc agc cgc act cgc act gtt ggt gag gtt gtc aac gcc atg 240
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met
65 70 75 80

aag gcc gag atc gct ggc 258
Lys Ala Glu Ile Ala Gly
85

<210> 13
<211> 86
<212> PRT
<213> Schizochytrium sp.

<400> 13

Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1 5 10 15

Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
20 25 30

Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
35 40 45

Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
50 55 60

Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met
65 70 75 80

Lys Ala Glu Ile Ala Gly
85

<210> 14
<211> 5
<212> PRT
<213> Schizochytrium sp.

<400> 14

Leu Gly Ile Asp Ser
1 5

<210> 15
<211> 21
<212> PRT
<213> Schizochytrium sp.

<400> 15

Ala Pro Ala Pro Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser
1 5 10 15

Ala Pro Ala Pro Ala
20

<210> 16
<211> 3006
<212> DNA
<213> Schizochytrium sp.

<400> 16
gcccccgccc cggtcaaggc tgctgcgcct gccgcccccg ttgcctcggc ccctgccccg 60

gctgtctcga acgagcttct tgagaaggcc gagactgtcg tcatggaggt cctcgccgcc 120

aagaccggct acgagaccga catgatcgag gctgacatgg agctcgagac cgagctcggc 180

attgactcca tcaagcgtgt cgagatcctc tccgaggtcc aggccatgct caatgtcgag 240

gccaaggatg tcgatgccct cagccgcact cgcactgttg gtgaggttgt caacgccatg 300

aaggccgaga tcgctggcag ctctgccccg gcgcctgctg ccgctgctcc ggctccggcc 360

aaggctgccc ctgccgccgc tgcgcctgct gtctcgaacg agcttctcga gaaggccgag 420

accgtcgtca tggaggtcct cgccgccaag actggctacg agactgacat gatcgagtcc 480

gacatggagc tcgagactga gctcggcatt gactccatca agcgtgtcga gatcctctcc 540

gaggttcagg ccatgctcaa cgtcgaggcc aaggacgtcg acgctctcag ccgcactcgc 600

actgtgggtg aggtcgtcaa cgccatgaag gctgagatcg ctggtggctc tgccccggcg 660

```
cctgccgccg ctgccccagg tccggctgct gccgcccctg cgcctgccgc cgccgcccct    720
gctgtctcga acgagcttct tgagaaggcc gagaccgtcg tcatggaggt cctcgccgcc    780
aagactggct acgagactga catgatcgag tccgacatgg agctcgagac cgagctcggc    840
attgactcca tcaagcgtgt cgagattctc tccgaggtcc aggccatgct caacgtcgag    900
gccaaggacg tcgacgctct cagccgcacc cgcactgttg gcgaggtcgt cgatgccatg    960
aaggccgaga tcgctggtgg ctctgccccg gcgcctgccg ccgctgctcc tgctccggct   1020
gctgccgccc ctgcgcctgc cgcccctgcg cctgctgtct cgagcgagct tctcgagaag   1080
gccgagactg tcgtcatgga ggtcctcgcc gccaagactg gctacgagac tgacatgatc   1140
gagtccgaca tggagctcga gaccgagctc ggcattgact ccatcaagcg tgtcgagatt   1200
ctctccgagg tccaggccat gctcaacgtc gaggccaagg acgtcgacgc tctcagccgc   1260
acccgcactg ttggcgaggt cgtcgatgcc atgaaggccg agatcgctgg tggctctgcc   1320
ccggcgcctg ccgccgctgc tcctgctccg gctgctgccg cccctgcgcc tgccgcccct   1380
gcgcctgccg cccctgcgcc tgctgtctcg agcgagcttc tcgagaaggc cgagactgtc   1440
gtcatggagg tcctcgccgc caagactggc tacgagactg acatgattga gtccgacatg   1500
gagctcgaga ccgagctcgg cattgactcc atcaagcgtg tcgagattct ctccgaggtt   1560
caggccatgc tcaacgtcga ggccaaggac gtcgacgctc tcagccgcac tcgcactgtt   1620
ggtgaggtcg tcgatgccat gaaggctgag atcgctggca gctccgcctc ggcgcctgcc   1680
gccgctgctc ctgctccggc tgctgccgct cctgcgcccg ctgccgccgc ccctgctgtc   1740
tcgaacgagc ttctcgagaa agccgagact gtcgtcatgg aggtcctcgc cgccaagact   1800
ggctacgaga ctgacatgat cgagtccgac atggagctcg agactgagct cggcattgac   1860
tccatcaagc gtgtcgagat cctctccgag gttcaggcca tgctcaacgt cgaggccaag   1920
gacgtcgatg ccctcagccg cacccgcact gttggcgagg ttgtcgatgc catgaaggcc   1980
gagatcgctg gtggctctgc cccggcgcct gccgccgctg cccctgctcc ggctgccgcc   2040
gcccctgctg tctcgaacga gcttctcgag aaggccgaga ctgtcgtcat ggaggtcctc   2100
gccgccaaga ctggctacga gaccgacatg atcgagtccg acatggagct cgagaccgag   2160
ctcggcattg actccatcaa gcgtgtcgag attctctccg aggttcaggc catgctcaac   2220
gtcgaggcca aggacgtcga tgctctcagc cgcactcgca ctgttggcga ggtcgtcgat   2280
gccatgaagg ctgagatcgc cggcagctcc gccccggcgc ctgccgccgc tgctcctgct   2340
ccggctgctg ccgctcctgc gcccgctgcc gctgcccctg ctgtctcgag cgagcttctc   2400
gagaaggccg agaccgtcgt catggaggtc ctcgccgcca agactggcta cgagactgac   2460
atgattgagt ccgacatgga gctcgagact gagctcggca ttgactccat caagcgtgtc   2520
gagatcctct ccgaggttca ggccatgctc aacgtcgagg ccaaggacgt cgatgccctc   2580
agccgcaccc gcactgttgg cgaggttgtc gatgccatga aggccgagat cgctggtggc   2640
tctgccccgg cgcctgccgc cgctgcccct gctccggctg ccgccgcccc tgctgtctcg   2700
```

aacgagcttc ttgagaaggc cgagaccgtc gtcatggagg tcctcgccgc caagactggc 2760

tacgagaccg acatgatcga gtccgacatg gagctcgaga ccgagctcgg cattgactcc 2820

atcaagcgtg tcgagattct ctccgaggtt caggccatgc tcaacgtcga ggccaaggac 2880

gtcgacgctc tcagccgcac tcgcactgtt ggcgaggtcg tcgatgccat gaaggctgag 2940

atcgctggtg gctctgcccc ggcgcctgcc gccgctgctc ctgcctcggc tggcgccgcg 3000

cctgcg 3006

<210> 17
<211> 2133
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(2133)

<400> 17

```
ttt ggc gct ctc ggc ggc ttc atc tcg cag cag gcg gag cgc ttc gag       48
Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
1               5                   10                  15

ccc gcc gaa atc ctc ggc ttc acg ctc atg tgc gcc aag ttc gcc aag       96
Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
            20                  25                  30

gct tcc ctc tgc acg gct gtg gct ggc ggc cgc ccg gcc ttt atc ggt      144
Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
        35                  40                  45

gtg gcg cgc ctt gac ggc cgc ctc gga ttc act tcg cag ggc act tct      192
Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
    50                  55                  60

gac gcg ctc aag cgt gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag      240
Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
65                  70                  75                  80

acc atc ggc ctc gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg      288
Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
                85                  90                  95

gac att gct cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg      336
Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
            100                 105                 110

cgc gag atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc      384
Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
        115                 120                 125

gca aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc      432
Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
    130                 135                 140

aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct ggc      480
Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly
145                 150                 155                 160

ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg cgc cag      528
Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln
                165                 170                 175
```

```
atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag gtc tct gcg      576
Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala
            180                 185                 190

agc gaa ccg gca tgg tgc gct ggc atc act gac gag aag gct gtg caa      624
Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln
        195                 200                 205

aag gct gct acc cag gag ctc aag cgc gcc ttt agc gct ggc gag ggc      672
Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly
    210                 215                 220

ccc aag ccc acg ccc cgc gct gtc act aag ctt gtg ggc tct gtt ctt      720
Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu
225                 230                 235                 240

ggc gct cgc gag gtg cgc agc tct att gct gcg att gaa gcg ctc ggc      768
Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly
                245                 250                 255

ggc aag gcc atc tac tcg tcg tgc gac gtg aac tct gcc gcc gac gtg      816
Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val
            260                 265                 270

gcc aag gcc gtg cgc gat gcc gag tcc cag ctc ggt gcc cgc gtc tcg      864
Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser
        275                 280                 285

ggc atc gtt cat gcc tcg ggc gtg ctc cgc gac cgt ctc atc gag aag      912
Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys
    290                 295                 300

aag ctc ccc gac gag ttc gac gcc gtc ttt ggc acc aag gtc acc ggt      960
Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly
305                 310                 315                 320

ctc gag aac ctc ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg     1008
Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met
                325                 330                 335

gtc ctc ttc agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct     1056
Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser
            340                 345                 350

gac tac gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc     1104
Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu
        355                 360                 365

gcc aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt     1152
Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
    370                 375                 380

ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc gtg     1200
Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val
385                 390                 395                 400

cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc atc gtg     1248
Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val
                405                 410                 415

ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg cgc acc ccg     1296
Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro
            420                 425                 430

tcc aag aag gtc ggc tcg gac acc atc acc ctg cac cgc aag att tcc     1344
Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser
        435                 440                 445
```

```
gcc aag tcc aac ccc ttc ctc gag gac cac gtc atc cag ggc cgc cgc     1392
Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg
    450                 455                 460

gtg ctg ccc atg acg ctg gcc att ggc tcg ctc gcg gag acc tgc ctc     1440
Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
465                 470                 475                 480

ggc ctc ttc ccc ggc tac tcg ctc tgg gcc att gac gac gcc cag ctc     1488
Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
                485                 490                 495

ttc aag ggt gtc act gtc gac ggc gac gtc aac tgc gag gtg acc ctc     1536
Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
            500                 505                 510

acc ccg tcg acg gcg ccc tcg ggc cgc gtc aac gtc cag gcc acg ctc     1584
Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
        515                 520                 525

aag acc ttt tcc agc ggc aag ctg gtc ccg gcc tac cgc gcc gtc atc     1632
Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
    530                 535                 540

gtg ctc tcc aac cag ggc gcg ccc ccg gcc aac gcc acc atg cag ccg     1680
Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
545                 550                 555                 560

ccc tcg ctc gat gcc gat ccg gcg ctc cag ggc tcc gtc tac gac ggc     1728
Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
                565                 570                 575

aag acc ctc ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc     1776
Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
            580                 585                 590

tcg tgc acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc     1824
Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
        595                 600                 605

tcc gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac     1872
Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
    610                 615                 620

ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc     1920
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
625                 630                 635                 640

cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc     1968
Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
                645                 650                 655

cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc     2016
Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
            660                 665                 670

tcc aac cag tcg ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc     2064
Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
        675                 680                 685

cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tcg gtc atc     2112
His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
    690                 695                 700

gcc acg gac agc ctt gcc ttc                                        2133
Ala Thr Asp Ser Leu Ala Phe
705                 710
```

<210> 18
<211> 711
<212> PRT
<213> Schizochytrium sp.

<400> 18

```
Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
1               5                   10                  15

Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
            20                  25                  30

Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
        35                  40                  45

Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
    50                  55                  60

Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
65                  70                  75                  80

Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
                85                  90                  95

Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
            100                 105                 110

Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
        115                 120                 125

Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
    130                 135                 140

Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly
145                 150                 155                 160

Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln
                165                 170                 175

Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala
            180                 185                 190

Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln
        195                 200                 205

Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly
    210                 215                 220

Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu
225                 230                 235                 240
```

```
Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly
                245                 250                 255

Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val
            260                 265                 270

Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser
        275                 280                 285

Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys
    290                 295                 300

Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly
305                 310                 315                 320

Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met
                325                 330                 335

Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser
            340                 345                 350

Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu
        355                 360                 365

Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
    370                 375                 380

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val
385                 390                 395                 400

Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val
                405                 410                 415

Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro
            420                 425                 430

Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser
        435                 440                 445

Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg
    450                 455                 460

Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
465                 470                 475                 480

Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
                485                 490                 495

Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
            500                 505                 510
```

```
Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
        515                 520                 525

Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
    530                 535                 540

Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
545                 550                 555                 560

Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
                565                 570                 575

Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
            580                 585                 590

Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
        595                 600                 605

Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
    610                 615                 620

Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
625                 630                 635                 640

Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
                645                 650                 655

Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
            660                 665                 670

Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
        675                 680                 685

His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
    690                 695                 700

Ala Thr Asp Ser Leu Ala Phe
705                 710


<210>  19
<211>  1350
<212>  DNA
<213>  Schizochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1350)

<400>  19
atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag       48
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1               5                   10                  15

cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc       96
```

```
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
            20                  25                  30

aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag      144
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
        35                  40                  45

gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac      192
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
    50                  55                  60

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac      240
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65                  70                  75                  80

ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc      288
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
                85                  90                  95

aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca      336
Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
            100                 105                 110

cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc      384
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
        115                 120                 125

cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc      432
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
    130                 135                 140

ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag      480
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145                 150                 155                 160

tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc      528
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
                165                 170                 175

tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc      576
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
            180                 185                 190

gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat      624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
        195                 200                 205

cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc      672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
    210                 215                 220

ctg ccg gag ccc ttt ttc atc ctt tcg ggc ttt tcc acc ttc cag gcc      720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225                 230                 235                 240

atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac      768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
                245                 250                 255

agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag      816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
            260                 265                 270

cgt ctc gat gat gcc atc cgc gac ggc gac cac atc tac ggc acc ctt      864
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
        275                 280                 285

ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc      912
```

```
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
    290                 295                 300

ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att      960
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305                 310                 315                 320

aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc     1008
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
                325                 330                 335

acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt     1056
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
            340                 345                 350

gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac     1104
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
        355                 360                 365

acc ctc gtc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc     1152
Thr Leu Val Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
    370                 375                 380

atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc     1200
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385                 390                 395                 400

aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag     1248
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
                405                 410                 415

acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt     1296
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
            420                 425                 430

ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc     1344
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
        435                 440                 445

gcc tgc                                                            1350
Ala Cys
    450
```

<210> 20
<211> 450
<212> PRT
<213> Schizochytrium sp.

<400> 20

```
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1               5                   10                  15

Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
            20                  25                  30

Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
        35                  40                  45

Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
    50                  55                  60

Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
```

```
65                  70                  75                  80

Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
                85                  90                  95

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
            100                 105                 110

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
        115                 120                 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
    130                 135                 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145                 150                 155                 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
                165                 170                 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
            180                 185                 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
        195                 200                 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
    210                 215                 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225                 230                 235                 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
                245                 250                 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
            260                 265                 270

Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
        275                 280                 285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
    290                 295                 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305                 310                 315                 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
                325                 330                 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
```

340 345 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
355 360 365

Thr Leu Val Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
370 375 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385 390 395 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
405 410 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
420 425 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
435 440 445

Ala Cys
450

<210> 21
<211> 1323
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1323)

<400> 21
tcg gcc cgc tgc ggc ggt gaa agc aac atg cgc atc gcc atc act ggt 48
Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly
1 5 10 15

atg gac gcc acc ttt ggc gct ctc aag gga ctc gac gcc ttc gag cgc 96
Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg
20 25 30

gcc att tac acc ggc gct cac ggt gcc atc cca ctc cca gaa aag cgc 144
Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg
35 40 45

tgg cgc ttt ctc ggc aag gac aag gac ttt ctt gac ctc tgc ggc gtc 192
Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val
50 55 60

aag gcc acc ccg cac ggc tgc tac att gaa gat gtt gag gtc gac ttc 240
Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe
65 70 75 80

cag cgc ctc cgc acg ccc atg acc cct gaa gac atg ctc ctc cct cag 288
Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln
85 90 95

cag ctt ctg gcc gtc acc acc att gac cgc gcc atc ctc gac tcg gga 336
Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly

```
            100                 105                 110

atg aaa aag ggt ggc aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac      384
Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp
        115                 120                 125

ctc gag ctc tac cgt cac cgt gct cgc gtc gct ctc aag gag cgc gtc      432
Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val
    130                 135                 140

cgc cct gaa gcc tcc aag aag ctc aat gac atg atg cag tac att aac      480
Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn
145                 150                 155                 160

gac tgc ggc aca tcc aca tcg tac acc tcg tac att ggc aac ctc gtc      528
Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val
                165                 170                 175

gcc acg cgc gtc tcg tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg      576
Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr
            180                 185                 190

atc acc gag ggc aac aac tcc gtc tac cgc tgc gcc gag ctc ggc aag      624
Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys
        195                 200                 205

tac ctc ctc gag acc ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc      672
Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val
    210                 215                 220

gat ctc tgc ggc agt gcc gaa aac ctt tac gtc aag tct cgc cgc ttc      720
Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe
225                 230                 235                 240

aag gtg tcc acc tcc gat acc ccg cgc gcc agc ttt gac gcc gcc gcc      768
Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala
                245                 250                 255

gat ggc tac ttt gtc ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt      816
Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg
            260                 265                 270

gag act agc tgc acc aag gac gac cgt atc tac gct tgc atg gat gcc      864
Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala
        275                 280                 285

atc gtc cct ggc aac gtc cct agc gcc tgc ttg cgc gag gcc ctc gac      912
Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp
    290                 295                 300

cag gcg cgc gtc aag ccg ggc gat atc gag atg ctc gag ctc agc gcc      960
Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala
305                 310                 315                 320

gac tcc gcc cgc cac ctc aag gac ccg tcc gtc ctg ccc aag gag ctc     1008
Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu
                325                 330                 335

act gcc gag gag gaa atc ggc ggc ctt cag acg atc ctt cgt gac gat     1056
Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp
            340                 345                 350

gac aag ctc ccg cgc aac gtc gca acg ggc agt gtc aag gcc acc gtc     1104
Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
        355                 360                 365

ggt gac acc ggt tat gcc tct ggt gct gcc agc ctc atc aag gct gcg     1152
Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
```

```
    370                 375                 380

ctt tgc atc tac aac cgc tac ctg ccc agc aac ggc gac gac tgg gat     1200
Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
385                 390                 395                 400

gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg tgc cag     1248
Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
                405                 410                 415

acc tcg cgc gct tgg ctc aag aac cct ggc gag cgt cgc tat gcg gcc     1296
Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
            420                 425                 430

gtc tcg ggc gtc tcc gag acg cgc tcg                                 1323
Val Ser Gly Val Ser Glu Thr Arg Ser
        435                 440


<210>  22
<211>  441
<212>  PRT
<213>  Schizochytrium sp.

<400>  22

Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly
1               5                   10                  15


Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg
            20                  25                  30


Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg
        35                  40                  45


Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val
    50                  55                  60


Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe
65                  70                  75                  80


Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln
                85                  90                  95


Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly
            100                 105                 110


Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp
        115                 120                 125


Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val
    130                 135                 140


Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn
145                 150                 155                 160


Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val
                165                 170                 175
```

```
Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr
            180                 185                 190

Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys
        195                 200                 205

Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val
    210                 215                 220

Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe
225                 230                 235                 240

Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala
                245                 250                 255

Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg
            260                 265                 270

Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala
        275                 280                 285

Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp
    290                 295                 300

Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala
305                 310                 315                 320

Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu
                325                 330                 335

Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp
            340                 345                 350

Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
        355                 360                 365

Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
    370                 375                 380

Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
385                 390                 395                 400

Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
                405                 410                 415

Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
            420                 425                 430

Val Ser Gly Val Ser Glu Thr Arg Ser
        435                 440
```

<210> 23
<211> 1500
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 23

```
tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac tac gag cgc gag       48
Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
1               5                  10                  15

aac cgc atc tcg ctc gac gag gag gcg ccc aag ctc att gtg ctt cgc       96
Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
            20                  25                  30

gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac aag atc cgc gag      144
Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
        35                  40                  45

cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag tcc gag ctc aag      192
Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
    50                  55                  60

gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc gag acc ctt gcc      240
Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
65                  70                  75                  80

cag gat gcc gct tct tca ggc tcg caa aag ccc ctc gct ctc agc ctc      288
Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
                85                  90                  95

gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag ctc gcg gcc aag      336
Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
            100                 105                 110

ggt atc ccg cgc tgc ctc aag atg cgc cgc gat tgg agc tcc cct gct      384
Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
        115                 120                 125

ggc agc cgc tac gcg cct gag ccg ctc gcc agc gac cgc gtc gcc ttc      432
Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
    130                 135                 140

atg tac ggc gaa ggt cgc agc cct tac tac ggc atc acc caa gac att      480
Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
145                 150                 155                 160

cac cgc att tgg ccc gaa ctc cac gag gtc atc aac gaa aag acg aac      528
His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn
                165                 170                 175

cgt ctc tgg gcc gaa ggc gac cgc tgg gtc atg ccg cgc gcc agc ttc      576
Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe
            180                 185                 190

aag tcg gag ctc gag agc cag cag caa gag ttt gat cgc aac atg att      624
Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile
        195                 200                 205

gaa atg ttc cgt ctt gga atc ctc acc tca att gcc ttc acc aat ctg      672
Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu
    210                 215                 220
```

```
gcg cgc gac gtt ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt      720
Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser
225                 230                 235                 240

ctt ggc gag att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc      768
Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu
                245                 250                 255

atc tcc gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac      816
Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn
            260                 265                 270

aag gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att      864
Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
        275                 280                 285

cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg cgc      912
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg
    290                 295                 300

ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc aag tac      960
Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr
305                 310                 315                 320

gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc att agc ggc     1008
Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly
                325                 330                 335

aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc ggt ggc aac att     1056
Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile
            340                 345                 350

cct gcg ctt ccc gtg acc cag ggc atg tgc ggc cac tgc ccc gag gtg     1104
Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val
        355                 360                 365

gga cct tat acc aag gat atc gcc aag atc cat gcc aac ctt gag ttc     1152
Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe
    370                 375                 380

ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc     1200
Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg
385                 390                 395                 400

ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc     1248
Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser
                405                 410                 415

ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc     1296
Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro
            420                 425                 430

caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag     1344
Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu
        435                 440                 445

gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt     1392
Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly
    450                 455                 460

ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat     1440
Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp
465                 470                 475                 480

gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt     1488
Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
                485                 490                 495
```

gtt cct ggc gtc 1500
Val Pro Gly Val
500

<210> 24
<211> 500
<212> PRT
<213> Schizochytrium sp.

<400> 24

Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
1 5 10 15

Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
20 25 30

Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
35 40 45

Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
50 55 60

Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
65 70 75 80

Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
85 90 95

Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
100 105 110

Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
115 120 125

Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
130 135 140

Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
145 150 155 160

His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn
165 170 175

Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe
180 185 190

Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile
195 200 205

Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu
210 215 220

```
Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser
225                 230                 235                 240

Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu
                245                 250                 255

Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn
            260                 265                 270

Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
        275                 280                 285

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg
    290                 295                 300

Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr
305                 310                 315                 320

Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly
                325                 330                 335

Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile
            340                 345                 350

Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val
        355                 360                 365

Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe
    370                 375                 380

Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg
385                 390                 395                 400

Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser
                405                 410                 415

Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro
            420                 425                 430

Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu
        435                 440                 445

Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly
    450                 455                 460

Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp
465                 470                 475                 480

Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
                485                 490                 495
```

```
Val Pro Gly Val
            500


<210>  25
<211>  1530
<212>  DNA
<213>  Schizochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1530)

<400>  25
ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt gcc tcc       48
Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
1               5                  10                  15

ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc att gtg ccg       96
Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
            20                  25                  30

ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc atg aaa acg tac      144
Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
        35                  40                  45

ggt gtt tcg gcg cct ctg tac acg ggc gcc atg gcc aag ggc att gcc      192
Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
    50                  55                  60

tct gcg gac ctc gtc att gcc gcc ggc cgc cag ggc atc ctt gcg tcc      240
Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
65                  70                  75                  80

ttt ggc gcc ggc gga ctt ccc atg cag gtt gtg cgt gag tcc atc gaa      288
Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
                85                  90                  95

aag att cag gcc gcc ctg ccc aat ggc ccg tac gct gtc aac ctt atc      336
Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
            100                 105                 110

cat tct ccc ttt gac agc aac ctc gaa aag ggc aat gtc gat ctc ttc      384
His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
        115                 120                 125

ctc gag aag ggt gtc acc ttt gtc gag gcc tcg gcc ttt atg acg ctc      432
Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu
    130                 135                 140

acc ccg cag gtc gtg cgg tac cgc gcg gct ggc ctc acg cgc aac gcc      480
Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala
145                 150                 155                 160

gac ggc tcg gtc aac atc cgc aac cgt atc att ggc aag gtc tcg cgc      528
Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg
                165                 170                 175

acc gag ctc gcc gag atg ttc atg cgt cct gcg ccc gag cac ctt ctt      576
Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu
            180                 185                 190

cag aag ctc att gct tcc ggc gag atc aac cag gag cag gcc gag ctc      624
Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu
        195                 200                 205
```

```
gcc cgc cgt gtt ccc gtc gct gac gac atc gcg gtc gaa gct gac tcg      672
Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
    210                 215                 220

ggt ggc cac acc gac aac cgc ccc atc cac gtc att ctg ccc ctc atc      720
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile
225                 230                 235                 240

atc aac ctt cgc gac cgc ctt cac cgc gag tgc ggc tac ccg gcc aac      768
Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
                245                 250                 255

ctt cgc gtc cgt gtg ggc gcc ggc ggt ggc att ggg tgc ccc cag gcg      816
Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
            260                 265                 270

gcg ctg gcc acc ttc aac atg ggt gcc tcc ttt att gtc acc ggc acc      864
Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
        275                 280                 285

gtg aac cag gtc gcc aag cag tcg ggc acg tgc gac aat gtg cgc aag      912
Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys
    290                 295                 300

cag ctc gcg aag gcc act tac tcg gac gta tgc atg gcc ccg gct gcc      960
Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala
305                 310                 315                 320

gac atg ttc gag gaa ggc gtc aag ctt cag gtc ctc aag aag gga acc     1008
Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr
                325                 330                 335

atg ttt ccc tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac     1056
Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr
            340                 345                 350

gac tcg ttc gag tcc atg ccc ccc gca gag ctt gcg cgc gtc gag aag     1104
Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys
        355                 360                 365

cgc atc ttc agc cgc gcg ctc gaa gag gtc tgg gac gag acc aaa aac     1152
Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn
    370                 375                 380

ttt tac att aac cgt ctt cac aac ccg gag aag atc cag cgc gcc gag     1200
Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu
385                 390                 395                 400

cgc gac ccc aag ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc     1248
Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
                405                 410                 415

ctg gcg agc cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac     1296
Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp
            420                 425                 430

tac cag gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc     1344
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile
        435                 440                 445

aag gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc     1392
Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
    450                 455                 460

gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc cgt     1440
Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg
465                 470                 475                 480
```

```
ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc gct ctt     1488
Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu
                485                 490                 495

gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag ctg             1530
Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
            500                 505                 510
```

<210> 26
<211> 510
<212> PRT
<213> Schizochytrium sp.

<400> 26

```
Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
1               5                   10                  15

Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
            20                  25                  30

Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
        35                  40                  45

Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
    50                  55                  60

Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
65                  70                  75                  80

Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
                85                  90                  95

Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
            100                 105                 110

His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
        115                 120                 125

Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu
    130                 135                 140

Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala
145                 150                 155                 160

Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg
                165                 170                 175

Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu
            180                 185                 190

Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu
        195                 200                 205
```

```
Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
    210                 215                 220

Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile
225                 230                 235                 240

Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
                245                 250                 255

Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
            260                 265                 270

Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
        275                 280                 285

Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys
    290                 295                 300

Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala
305                 310                 315                 320

Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr
                325                 330                 335

Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr
            340                 345                 350

Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys
        355                 360                 365

Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn
    370                 375                 380

Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu
385                 390                 395                 400

Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
                405                 410                 415

Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp
            420                 425                 430

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile
        435                 440                 445

Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
    450                 455                 460

Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg
465                 470                 475                 480
```

```
Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu
                485                 490                 495


Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
            500                 505                 510


<210>  27
<211>  1350
<212>  DNA
<213>  Schizochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1350)

<400>  27
atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc       48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1               5                   10                  15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa       96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
            20                  25                  30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag      144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
        35                  40                  45

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc      192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac      240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc      288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc      336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac      384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc      432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att      480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc      528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg      576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc      624
```

```
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc      672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag      720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg      768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
                245                 250                 255

gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa      816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
            260                 265                 270

ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac      864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
        275                 280                 285

cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc      912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
    290                 295                 300

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag      960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305                 310                 315                 320

gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag     1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
                325                 330                 335

atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac     1056
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
            340                 345                 350

ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa     1104
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag     1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
    370                 375                 380

atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac     1200
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385                 390                 395                 400

atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc     1248
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
                405                 410                 415

atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac     1296
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
            420                 425                 430

ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac     1344
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
        435                 440                 445

ccc tcc                                                            1350
Pro Ser
    450
```

<210> 28
<211> 450
<212> PRT
<213> Schizochytrium sp.

<400> 28

```
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1               5                  10                  15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
            20                  25                  30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
        35                  40                  45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
```

245 250 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
260 265 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
275 280 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
290 295 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
325 330 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
340 345 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
370 375 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385 390 395 400

Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
405 410 415

Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
420 425 430

Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
435 440 445

Pro Ser
450

<210> 29
<211> 1497
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1497)

<400> 29
aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc ccc gag 48
Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu

```
1               5                   10                  15

gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc gcc ccg       96
Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
            20                  25                  30

gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc gtt gcc      144
Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
        35                  40                  45

ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag gag atg      192
Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
    50                  55                  60

tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc tct ttt      240
Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
65                  70                  75                  80

gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc ttc ccc      288
Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
                85                  90                  95

ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg ctc acc      336
Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
            100                 105                 110

tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg tgc ctc      384
Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
        115                 120                 125

ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc agc ccc      432
Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
    130                 135                 140

gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct gac ctc      480
Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
145                 150                 155                 160

aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag ggt acc      528
Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr
                165                 170                 175

atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac aag ggc      576
Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly
            180                 185                 190

gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag atc gcc      624
Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala
        195                 200                 205

ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc ctg acc      672
Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr
    210                 215                 220

atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac gcc gag      720
Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu
225                 230                 235                 240

ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc aac gtc      768
Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val
                245                 250                 255

acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc cac cgc      816
Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg
            260                 265                 270

ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag ggc tcg      864
Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser
```

```
        275                 280                 285

acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag gcc ggc      912
Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly
    290                 295                 300

ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac aag gtt      960
Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val
305                 310                 315                 320

ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc agc ggc     1008
Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly
                325                 330                 335

cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc aac cgc     1056
Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg
            340                 345                 350

cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc tcc ggc     1104
Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly
        355                 360                 365

agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg gtc aac     1152
Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn
    370                 375                 380

ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg gtc atg     1200
Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met
385                 390                 395                 400

ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag gcc atc     1248
Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile
                405                 410                 415

gcc gcc cac gag gat ctc gct ggc aag cac ggc att gcc aac ccc acc     1296
Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn Pro Thr
            420                 425                 430

ttt gtg cac gcc ccg ggc aag atc agc tgg aag tac cgc ggc cag ctc     1344
Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly Gln Leu
        435                 440                 445

acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg tcc gtg     1392
Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser Val
    450                 455                 460

gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc tgg     1440
Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu Trp
465                 470                 475                 480

gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc atc     1488
Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg Ile
                485                 490                 495

gcc tcc ggt                                                        1497
Ala Ser Gly
```

<210> 30
<211> 499
<212> PRT
<213> Schizochytrium sp.

<400> 30

```
Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
1               5                   10                  15
```

```
Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
            20                  25                  30

Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
        35                  40                  45

Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
    50                  55                  60

Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
65                  70                  75                  80

Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
                85                  90                  95

Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
            100                 105                 110

Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
        115                 120                 125

Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
    130                 135                 140

Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
145                 150                 155                 160

Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr
                165                 170                 175

Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly
            180                 185                 190

Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala
        195                 200                 205

Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr
    210                 215                 220

Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu
225                 230                 235                 240

Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val
                245                 250                 255

Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg
            260                 265                 270

Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser
        275                 280                 285
```

```
Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly
    290                 295                 300

Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val
305                 310                 315                 320

Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly
                325                 330                 335

Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg
            340                 345                 350

Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly
        355                 360                 365

Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn
    370                 375                 380

Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met
385                 390                 395                 400

Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile
                405                 410                 415

Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn Pro Thr
            420                 425                 430

Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly Gln Leu
        435                 440                 445

Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser Val
    450                 455                 460

Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu Trp
465                 470                 475                 480

Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg Ile
                485                 490                 495

Ala Ser Gly
```

<210> 31
<211> 1512
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1512)

<400> 31

```
gcc ccg ctc tac ctc tcg cag gac ccg acc agc ggc cag ctc aag aag       48
Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
1               5                   10                  15

cac acc gac gtg gcc tcc ggc cag gcc acc atc gtg cag ccc tgc acg       96
His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
            20                  25                  30

ctc ggc gac ctc ggt gac cgc tcc ttc atg gag acc tac ggc gtc gtc      144
Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
        35                  40                  45

gcc ccg ctg tac acg ggc gcc atg gcc aag ggc att gcc tcg gcg gac      192
Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
    50                  55                  60

ctc gtc atc gcc gcc ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc      240
Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
65                  70                  75                  80

ggc ggc ctc ccc atg cac cac gtg cgc gcc gcc ctc gag aag atc cag      288
Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
                85                  90                  95

gcc gcc ctg cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct      336
Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
            100                 105                 110

ttt gac agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag      384
Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
        115                 120                 125

ggc gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag      432
Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
    130                 135                 140

gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt tcg      480
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
145                 150                 155                 160

gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc gag ctc      528
Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
                165                 170                 175

gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc gag aag ctc      576
Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
            180                 185                 190

atc gcc tcg ggc gag atc acc cag gag cag gcc gag ctc gcg cgc cgc      624
Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg
        195                 200                 205

gtt ccc gtc gcc gac gat atc gct gtc gag gct gac tcg ggc ggc cac      672
Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
    210                 215                 220

acc gac aac cgc ccc atc cac gtc atc ctc ccg ctc atc atc aac ctc      720
Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
225                 230                 235                 240

cgc aac cgc ctg cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc      768
Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
                245                 250                 255

cgc gtt ggc gcc ggc ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc      816
Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
            260                 265                 270
```

```
gcg ctc acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag      864
Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
        275                 280                 285

gtc gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg      912
Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser
    290                 295                 300

cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg ttc      960
Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe
305                 310                 315                 320

gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg ttc ccc     1008
Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro
                325                 330                 335

tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc     1056
Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe
            340                 345                 350

gac tcc atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc     1104
Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe
        355                 360                 365

aag cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att     1152
Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
    370                 375                 380

aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc     1200
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
385                 390                 395                 400

aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc agc     1248
Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
                405                 410                 415

cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac cag gtc     1296
Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
            420                 425                 430

tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc aag ggc acc     1344
Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
        435                 440                 445

tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt gtc gtc cag atc     1392
Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
    450                 455                 460

aac ctg caa atc ctc cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc     1440
Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
465                 470                 475                 480

ctg cgc aac gac ccg cgc att gac ctc gag acc gag gat gct gcc ttt     1488
Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
                485                 490                 495

gtc tac gag ccc acc aac gcg ctc                                    1512
Val Tyr Glu Pro Thr Asn Ala Leu
            500
```

<210> 32
<211> 504
<212> PRT
<213> Schizochytrium sp.

<400> 32

```
Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
1               5                  10                  15

His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
            20                  25                  30

Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
        35                  40                  45

Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
    50                  55                  60

Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
65                  70                  75                  80

Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
                85                  90                  95

Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
            100                 105                 110

Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
        115                 120                 125

Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
    130                 135                 140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
145                 150                 155                 160

Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
                165                 170                 175

Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
            180                 185                 190

Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg
        195                 200                 205

Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
    210                 215                 220

Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
225                 230                 235                 240

Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
                245                 250                 255

Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
            260                 265                 270
```

```
Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
        275                 280                 285

Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser
    290                 295                 300

Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe
305                 310                 315                 320

Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro
                325                 330                 335

Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe
            340                 345                 350

Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe
        355                 360                 365

Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
    370                 375                 380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
385                 390                 395                 400

Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
                405                 410                 415

Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
            420                 425                 430

Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
        435                 440                 445

Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
    450                 455                 460

Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
465                 470                 475                 480

Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
                485                 490                 495

Val Tyr Glu Pro Thr Asn Ala Leu
            500
```

<210> 33
<211> 714
<212> DNA
<213> Nostoc sp.

<400> 33
atgttgcagc atacttggct accaaaaccc ccaaatttaa ccttattgtc agatgaagtt 60

catctctggc gcattcccct tgaccaacca gaatcacagc tacaggattt agccgctacc 120

ttatctagtg acgaattagc ccgtgcaaac agattttatt ttcccgaaca tcgccggcgt 180

tttactgctg gtcgtggtat tctccgcagt atcttggggg gctatttggg tgtggaacca 240

gggcaagtta aatttgatta tgaatcccgt ggtaaaccaa tattaggcga tcgctttgcc 300

gagagtggtt tattatttaa cttgtcacac tcccagaact tggccttgtg tgcagtcaat 360

tacacgcgcc aaatcggcat cgatttagaa tatctccgcc ccacatctga tttagaatcc 420

cttgccaaaa ggttcttttt accgcgagaa tatgaattat tgcgatcgct acccgatgag 480

caaaaacaaa aaattttctt tcgttactgg acttgtaaag aggcttatct taaagcaacg 540

ggtgacggca tcgctaaatt agaggaaatt gaaatagcac taactcccac agaaccagct 600

aagttacaga cagctccagc gtggagtctc ctagagctag tgccagatga taattgtgtt 660

gctgctgttg ccgtggcggg ttttggctgg cagccaaaat tctggcatta ttga 714

<210> 34
<211> 237
<212> PRT
<213> Nostoc sp.

<400> 34

Met Leu Gln His Thr Trp Leu Pro Lys Pro Pro Asn Leu Thr Leu Leu
1 5 10 15

Ser Asp Glu Val His Leu Trp Arg Ile Pro Leu Asp Gln Pro Glu Ser
20 25 30

Gln Leu Gln Asp Leu Ala Ala Thr Leu Ser Ser Asp Glu Leu Ala Arg
35 40 45

Ala Asn Arg Phe Tyr Phe Pro Glu His Arg Arg Arg Phe Thr Ala Gly
50 55 60

Arg Gly Ile Leu Arg Ser Ile Leu Gly Gly Tyr Leu Gly Val Glu Pro
65 70 75 80

Gly Gln Val Lys Phe Asp Tyr Glu Ser Arg Gly Lys Pro Ile Leu Gly
85 90 95

Asp Arg Phe Ala Glu Ser Gly Leu Leu Phe Asn Leu Ser His Ser Gln
100 105 110

Asn Leu Ala Leu Cys Ala Val Asn Tyr Thr Arg Gln Ile Gly Ile Asp
115 120 125

Leu Glu Tyr Leu Arg Pro Thr Ser Asp Leu Glu Ser Leu Ala Lys Arg
130 135 140

Phe Phe Leu Pro Arg Glu Tyr Glu Leu Leu Arg Ser Leu Pro Asp Glu

145 150 155 160

Gln Lys Gln Lys Ile Phe Phe Arg Tyr Trp Thr Cys Lys Glu Ala Tyr
165 170 175

Leu Lys Ala Thr Gly Asp Gly Ile Ala Lys Leu Glu Glu Ile Glu Ile
180 185 190

Ala Leu Thr Pro Thr Glu Pro Ala Lys Leu Gln Thr Ala Pro Ala Trp
195 200 205

Ser Leu Leu Glu Leu Val Pro Asp Asp Asn Cys Val Ala Ala Val Ala
210 215 220

Val Ala Gly Phe Gly Trp Gln Pro Lys Phe Trp His Tyr
225 230 235

<210> 35
<211> 8733
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 35
atggctgcta ggttgcaaga acaaaaaggt ggtgagatgg atactagaat tgctatcatt 60

ggaatgtctg ctattttgcc atgtggtact actgttagag aatcttggga aactattaga 120

gctggtattg attgtttgtc tgatttgcct gaagatagag ttgatgttac tgcttacttt 180

gatccagtta aaactactaa agataaaatc tattgtaaga gaggtggttt cattccagaa 240

tatgattttg atgctagaga atttggtttg aatatgtttc agatggaaga ttctgatgct 300

aatcaaacta tttctttgtt gaaagttaaa gaagcattgc aagatgctgg catcgatgct 360

ttgggtaaag agaagaagaa tattggttgt gttttgggta ttggtggtgg tcaaaaatct 420

tctcatgaat tttactcaag attgaattat gttgttgttg agaaggtatt gagaaaaatg 480

ggtatgccag aagaagatgt taaagttgct gttgaaaaat acaaagctaa ttttccagag 540

tggagattgg attcttttcc aggtttcttg ggaaatgtta ctgcaggaag atgtactaat 600

acttttaatc ttgatggcat gaattgtgtt gttgatgctg cttgtgcttc ttctttgatt 660

gctgttaaag ttgctattga tgaattgttg tacggtgatt gtgatatgat ggttactggt 720

gctacttgta ctgataattc tattggaatg tacatggctt tttctaaaac tccagttttc 780

tctactgatc catctgttag agcttatgat gaaaaaacta aaggaatgtt gattggtgaa 840

ggttctgcta tgttggtttt gaaaagatat gctgatgctg ttagagatgg tgatgaaatt 900

catgctgtta ttagaggttg tgcttcttct tctgatggta aagctgctgg tatctatact 960

ccaactattt ctggtcaaga agaagcattg agaagagctt ataatagagc ttgtgttgat 1020

ccagctactg ttactttggt tgaaggtcat ggtactggta ctccagttgg tgatagaatt 1080

gaattgactg ctttgagaaa tttgtttgat aaagcatatg gtgaaggtaa tactgaaaaa 1140

gttgctgttg gttctattaa atcttctatt ggtcatttga aagctgttgc tggtttggct 1200

ggaatgatta aagttatcat ggctttgaaa cataaaactt tgccaggaac tattaatgtt 1260

gataatccac caaacttgta cgataatact ccaattaacg aatcttcttt gtacattaat 1320

actatgaata gaccttggtt tccaccacca ggtgttccaa gaagagctgg tatttcttct 1380

tttggttttg gtggtgctaa ttatcatgct gttttggaag aagctgaacc agaacatact 1440

actgcttata ggttgaacaa aagaccacaa ccagttttga tgatggctgc tactccagct 1500

gctttgcaat ctttgtgtga agctcaattg aaagaatttg aagctgctat taaagaaaac 1560

gaaactgtta aaaatactgc ttatattaaa tgtgttaaat ttggtgaaca attcaaattc 1620

cctggtagta ttccagctac taatgctagg ttgggtttct tggttaaaga tgctgaagat 1680

gcttgttcta ctttgagagc tatttgtgct caatttgcta aagatgttac taaagaagca 1740

tggagattgc caagagaagg tgtttctttt agagctaaag gtattgctac taatggtgct 1800

gttgctgctt tgttttctgg tcaaggtgct caatatactc atatgttttc tgaagttgct 1860

atgaattggc cacaattcag acaatctatt gctgctatgg atgctgctca atctaaagtt 1920

gctggttctg ataaagattt tgaaagagtt tctcaagttt tgtatccaag aaaaccatac 1980

gagagagaac cagagcaaga tcataagaag atttctttga ctgcttattc tcaaccatct 2040

actttggctt gtgctttggg tgcttttgaa atttttaaag aagctggttt tactccagat 2100

tttgctgctg gtcattcttt gggtgaattt gctgctttgt acgctgctgg ttgtgttgat 2160

agagatgaat tgtttgaatt ggtttgtaga agagctagaa ttatgggtgg taaagatgct 2220

ccagctactc caaaaggttg catggctgct gttattggtc caaatgctga aaatattaaa 2280

gttcaagctg ctaatgtttg gttaggaaat tctaattctc catctcaaac tgttattact 2340

ggttctgttg aaggtattca agctgaatct gctaggttgc aaaaagaagg ttttagagtt 2400

gttccattgg cttgtgaatc tgcttttcat tctccacaga tggaaaatgc ttcttctgct 2460

tttaaagatg ttatctctaa agtttctttt agaactccaa aagctgaaac taaattgttt 2520

tctaatgttt ctggtgaaac ttatccaact gatgctagag aaatgttgac tcaacatatg 2580

acttcttctg ttaaattttt gactcaagtt agaaatatgc atcaagctgg tgctagaatt 2640

tttgttgaat tcggtccaaa acaagttttg tctaaattgg tttctgaaac tttgaaagat 2700

gatccatctg ttgttactgt ttctgttaat ccagcttctg gtactgattc tgatattcaa 2760

ttgagagatg ctgctgttca attggttgtt gctggtgtta atttgcaagg ttttgataaa 2820

tgggatgctc cagatgctac tagaatgcaa gctattaaaa aaaaaagaac tactttgaga 2880

ttgtctgctg ctacttatgt ttctgataaa actaagaaag ttagagatgc tgctatgaat 2940

gatggtagat gtgttactta cttgaaaggt gctgctccat tgattaaagc tccagaacca 3000

gttgttgatg aagctgctaa aagagaagct gaaagattgc aaaaagaatt gcaagatgct 3060

caaagacaat tggatgatgc taaaagagct gctgctgaag ctaattctaa attggctgct 3120

gctaaagaag aagctaaaac tgctgctgct tctgctaaac cagctgttga tactgctgtt 3180

gttgaaaaac atagagctat tttgaaatct atgttggctg aattggatgg ttatggttct 3240

gttgatgctt cttctttgca acaacaacaa caacaacaaa ctgctccagc tccagttaaa 3300

gctgctgctc cagctgctcc agttgcttct gctccagcac ccgcagttag caacgaactc 3360

ttagaaaaag ccgagacagt agtgatggaa gttcttgcag ctaaaacggg gtacgaaaca 3420

gatatgattg aagcagatat ggaacttgaa actgaactgg gcattgattc gattaaacgc 3480

gtggaaattc tgtcagaagt gcaagctatg ttaaatgttg aagcgaaaga tgttgatgca 3540

ctgtcacgca cacgcaccgt gggcgaagta gtgaacgcca tgaaagcaga aattgcaggc 3600

tcctcagcac ccgcgccggc cgcagcagca ccagcccccg caaaagccgc ccccgcagcg 3660

gcggctccag ccgtttcaaa cgaattactc gaaaaagcag aaaccgtagt gatggaagtc 3720

cttgccgcca aaacgggtta tgagaccgat atgatcgaaa gcgatatgga attagaaacc 3780

gaattaggga ttgatagtat taaacgcgta gaaattctgt ccgaagtaca agctatgctg 3840

aatgtagaag caaaagatgt agatgcgtta agccgcacac gcactgttgg tgaagttgtg 3900

aatgctatga aagctgaaat tgcaggaggt tcagcaccgg ccccagcagc cgcagcccca 3960

ggtccagcag cagccgcacc ggcccccgcc gccgccgcac cggcagtatc aaacgagttg 4020

ttagagaaag cggaaaccgt tgtgatggaa gtacttgccg cgaagacagg ttacgagacc 4080

gatatgatcg aaagtgacat ggaattagaa accgaattgg gcattgatag cattaaacgc 4140

gtagaaattt tatccgaagt tcaagccatg ttaaatgttg aagccaaaga tgtggatgcg 4200

ttatcccgca cgcgtaccgt cggagaagta gtggacgcta tgaaagcaga gattgcagga 4260

ggaagtgcac cggctccagc agcagcagca cccgccccag cggcagcggc gccggcaccg 4320

gccgctccgg ccccagccgt tagttcagaa ctcctcgaaa aagcagaaac tgttgtcatg 4380

gaagtattag ctgcaaaaac aggttacgag acggatatga ttgaaagcga tatggaatta 4440

gaaaccgaat taggcattga ttcaattaaa cgtgttgaaa tcttaagtga agtccaagcc 4500

atgcttaatg ttgaagccaa agatgtagat gcattatctc gcacgcgtac agtgggtgaa 4560

gttgtcgatg cgatgaaagc agaaatcgcg ggaggatcag cgccagcccc ggcagcagca 4620

gcccccgcgc ccgccgcggc cgcacctgcg ccggccgccc cagcccctgc agcaccggcc 4680

ccagcagtgt cgtcggaatt actcgaaaaa gctgaaacgg tcgttatgga agtacttgct 4740

gcaaagacgg gctatgaaac ggatatgatt gaatcggata tggaattaga aacagaactt 4800

ggtattgact ctattaaacg cgtggaaatt ctgagcgaag tacaggcaat gttaaacgta 4860

gaagccaaag atgtagacgc tttgtcacgc acacggacgg taggagaagt tgtggatgcg 4920

atgaaagctg aaattgccgg ttcaagtgct agcgcccctg ctgccgccgc ccctgcccct 4980

gccgccgcag caccggcccc ggcagccgca gctccagcag ttagtaacga attactcgaa 5040

aaagcagaaa cggtggtcat ggaagtgtta gcagcaaaaa ctggatatga aacggacatg 5100

attgaaagcg atatggaatt agaaacagaa ctgggaattg atagtattaa acgtgttgag 5160

attttatctg aggttcaagc tatgctgaat gttgaagcga aagatgtaga cgcactgtct 5220

cggacccgca cagtaggtga agtggtggac gcgatgaaag cagaaatcgc aggtggaagt 5280

gctccggccc cggcggcagc cgcacccgcg cccgcggccg cagccccagc agttagcaac 5340

gaattactcg agaaagcaga aactgtagtg atggaagtgt tagccgcaaa aacgggttat 5400

gaaacggata tgattgaaag cgatatggaa ctggaaaccg aactgggcat tgattctatt 5460

aaacgtgtcg aaatcttatc ggaagtccaa gcaatgctga acgtagaggc aaaggatgtt 5520

gatgccctgt cacgtacccg taccgtaggt gaagttgtag atgccatgaa agctgaaatc 5580

gcaggcagta gcgccccggc accagccgcc gccgcccccg cgccggcagc cgccgcaccc 5640

gcgccagccg cagctgctcc agctgtatct agtgagctgc tcgaaaaagc agaaaccgtg 5700

gttatggaag tgctcgccgc taaaacagga tatgaaaccg atatgattga aagcgatatg 5760

gaattagaaa ccgaactggg tattgatagt attaagcgtg ttgaaatttt gtcagaagtt 5820

caagctatgt tgaatgtaga agccaaagat gtagacgctt taagtcggac gcgtactgtt 5880

ggagaagtcg tagacgccat gaaagcagag attgcaggcg gaagtgcacc ggccccggca 5940

gcagcagccc cagcaccagc ggccgccgct cctgcagtgt caaacgaact tctggaaaaa 6000

gctgaaaccg tcgtcatgga agtgctggct gcaaaaactg gatatgaaac agacatgatt 6060

gaatcagata tggaactcga aaccgaactg gggattgata gcattaaacg tgtggaaatt 6120

ttatcggagg tacaagcaat gttaaatgtg gaagcaaaag atgtggatgc actgagccgt 6180

actcgtactg ttggtgaggt cgtggatgcg atgaaagcag aaattgctgg agggagtgcg 6240

cctgccccgg ccgccgccgc acccgcgtct gccggtgctg cccccgctgt caaaattgat 6300

tctgttcatg gtgctgattg tgatgatttg tctttgatgc atgctaaagt tgttgatatt 6360

agaagaccag atgaattgat tttggaaaga ccagaaaata gaccagtttt ggttgttgat 6420

gatggttctg aattgacttt ggctttggtt agagttttgg gtgcttgtgc tgttgttttg 6480

acttttgaag gtttgcaatt ggctcaaaga gctggtgctg ctgctattag acatgttttg 6540

gctaaagatt tgtctgctga atctgctgaa aaagctatta aagaagctga acaaagattt 6600

ggtgctttgg gtggttttat ctctcaacaa gctgaaagat ttgaaccagc tgaaattttg 6660

ggttttactt tgatgtgtgc taaatttgct aaagcatctt tgtgcactgc tgttgctggt 6720

ggtagaccag ctttcattgg tgttgctagg ttggatggta ggttgggttt tacttctcaa 6780

ggaacttctg atgctttgaa aagagctcaa agaggtgcta tttttggttt gtgcaagact 6840

attggtttgg aatggtctga atctgatgtt ttctcaagag gtgttgatat tgctcaaggt 6900

atgcatccag aagatgctgc tgttgctatt gttagagaaa tggcttgtgc tgatattaga 6960

attagagaag ttggtattgg tgctaatcaa caaagatgta ctattagagc tgctaaattg 7020

gaaactggaa atccacaaag acaaattgct aaagatgatg ttttgttggt ttctggtggt 7080

gctagaggaa ttactccatt gtgcattaga gaaattacta gacaaattgc tggtggaaag 7140

tatattttgt tgggtaggtc taaagtttct gcttctgaac cagcttggtg tgctggtatt 7200

actgatgaaa aagctgttca aaaagctgct actcaagaat tgaaaagagc tttttctgct 7260

ggtgaaggtc caaaaccaac tccaagagct gttactaaat tggttggttc tgttttgggt 7320

gctagagaag ttaggtcttc tattgctgct attgaagcat tgggtggaaa agctatctat 7380

tcttcttgtg atgttaattc tgctgctgat gttgctaaag ctgttagaga tgctgaatct 7440

caattgggtg ctagagtttc tggtattgtt catgcttctg gtgttttgag agataggttg 7500

attgaaaaaa aattgccaga tgaatttgat gctgtttttg gtactaaagt tactggtttg 7560

gaaaatttgt tggctgctgt tgatagagct aatttgaaac atatggtttt gttttcttct 7620

ttggctggtt ttcatggtaa tgttggtcaa tctgattatg ctatggctaa cgaagcattg 7680

aacaaaatgg gtttggaatt ggctaaagat gtttctgtta aatctatttg ttttggtcct 7740

tgggatggtg gtatggttac tccacaattg aaaaaacaat ttcaagaaat gggtgttcaa 7800

attattccaa gagaaggtgg tgctgatact gttgctagaa ttgttttggg ttcttctcca 7860

gctgaaattt tggttggtaa ttggagaact ccatctaaaa aagttggttc tgatactatt 7920

actttgcaca gaaaaatttc tgctaaatct aatccatttt tggaagatca tgtcattcaa 7980

ggtagaagag ttttgccaat gactttggct attggttctt tggctgaaac ttgtttgggt 8040

ttgtttcctg gatattcttt gtgggctatt gatgatgctc aattgtttaa aggtgttact 8100

gttgatggtg atgttaattg tgaagttact ttgactccat ctactgctcc ttctggtaga 8160

gttaatgttc aagctacttt gaaaactttt tcttctggta aattggttcc agcttataga 8220

gctgttattg ttttgtctaa tcaaggtgct ccaccagcta atgctactat gcaaccacca 8280

tctttggatg ctgatccagc tttgcaaggt tctgtttatg atggaaagac tttgtttcat 8340

ggtccagctt ttagaggtat tgatgatgtt ttgtcttgta ctaaatctca attggttgct 8400

aaatgttctg ctgttccagg ttctgatgct gctagaggtg aatttgctac tgatactgat 8460

gctcatgatc catttgttaa tgatttggct tttcaagcta tgttggtttg ggttagaaga 8520

actttgggtc aagctgcttt gccaaattct attcaaagaa ttgttcaaca cagaccagtt 8580

ccacaagata aaccatttta tattactttg agatctaatc aatctggtgg tcattctcaa 8640

cataaacatg ctttgcaatt tcataacgaa caaggtgatt tgttcattga tgttcaagca 8700

tctgttattg ctactgattc tttggctttt taa 8733

<210> 36
<211> 6180
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 36
atggctgcta gaaatgtttc tgctgctcat gaaatgcatg atgaaaaaag aattgctgtt 60

gttggtatgg ctgttcaata tgctggttgt aagactaaag atgaattttg ggaagttttg 120

atgaatggta aagttgaatc taaagttatc tctgataaaa gattgggttc taattaccga 180

gctgaacatt acaaggctga aagatccaaa tacgctgata ctttttgtaa cgaaacttat 240

ggtactttgg atgaaaacga aattgataac gaacatgaat tgttgttgaa tttggctaaa 300

caagcattgg ctgaaacttc tgttaaagat tctactagat gtggtattgt ttctggttgt 360

ttgtcttttc ctatggataa tttgcaaggt gaattgttga atgtctatca aaatcatgtt 420

gagaagaaat tgggtgctag agtttttaaa gatgcttctc attggtctga aagagaacaa 480

tctaacaaac cagaagctgg tgatagaaga attttcatgg acccagcttc ttttgttgct 540

gaagaattga atttgggtgc tttgcattat tctgttgatg ctgcttgtgc tactgcttta 600

tacgttttga gattggctca agatcatttg gtttctggtg ctgctgatgt tatgttgtgt 660

ggtgctactt gtttgccaga accattcttt atcttgtctg gtttttctac ttttcaagct 720

atgccagttg gtactggtca aaatgtttct atgccattgc ataaagattc tcaaggtttg 780

actccaggtg aaggtggttc tatcatggtt ttgaaaagat tggatgatgc tattagagat 840

ggtgatcata tctatggtac tttgttgggt gctaatgttt ctaattctgg cactggtttg 900

ccattgaaac cattgttgcc atctgaaaaa aaatgtttga tggatactta tactagaatt 960

aatgttcatc cacataaaat tcaatatgtt gaatgtcatg ctactggtac tccacaaggt 1020

gatagggttg aaattgatgc tgttaaagca tgttttgaag gaaaagttcc aagatttggt 1080

actactaaag gaaactttgg tcatactttg gttgctgctg gttttgctgg aatgtgcaaa 1140

gttttgttgt ctatgaaaca tggtatcatt ccaccaactc caggtattga tgatgaaact 1200

aagatggacc cattggttgt ttctggtgaa gctattcctt ggccagaaac taatggtgaa 1260

ccaaaaagag ctggtttgtc tgcttttggt tttggtggta ctaatgctca tgctgttttt 1320

gaagaacatg atccatctaa tgctgcttgt actggtcatg attctatttc tgctttgtct 1380

gctagatgtg gtggtgaatc taatatgaga attgctatta ctggtatgga tgctactttt 1440

ggtgctttga aaggtttgga tgcttttgaa agagccatct acactggtgc tcatggtgct 1500

attccattgc cagaaaagag atggagattt ttgggcaaag ataaagattt cttggatttg 1560

tgtggtgtta aagctactcc acatggttgt tatattgaag atgttgaagt tgattttcaa 1620

agattgagaa ctccaatgac tccagaagat atgttgttgc cacaacaatt gttggctgtt 1680

actactattg atagagctat tttggattct ggtatgaaaa aaggtggtaa tgttgctgtt 1740

tttgttggtt tgggtaccga tttggaattg tacagacata gagctagagt tgctttgaaa 1800

gaaagagtta gaccagaagc atctaaaaaa ttgaatgata tgatgcagta cattaatgat 1860

tgtggcacct ctacttctta tacttcttat attggtaatt tggttgctac tagagtttct 1920

tctcaatggg gttttactgg tccatctttt actattactg aagggaataa ctctgtttat 1980

agatgtgctg aattgggaaa gtatttgttg gaaactggtg aagttgatgg tgttgttgtt 2040

gctggtgttg atttgtgtgg ttctgctgaa aacttatacg ttaaatcaag aagattcaaa 2100

gtttctactt ctgatactcc aagagcttct tttgatgctg ctgctgatgg ttactttgtt 2160

ggtgaaggtt gtggtgcttt tgttttgaaa agagaaactt cttgtactaa agatgataga 2220

atctatgctt gcatggatgc tattgttcca ggtaatgttc catctgcttg tttgagagaa 2280

gcattggatc aagctagagt taaaccaggt gatattgaaa tgttggaatt gtctgctgat 2340

tctgctagac atttgaaaga tccatctgtt ttgccaaaag aattgactgc tgaagaagaa 2400

attggtggtt tgcaaactat tttgagagat gatgataaat tgccaagaaa tgttgctact 2460

ggttctgtta aagctactgt tggtgatact ggttatgctt ctggtgctgc ttctttgatt 2520

aaagctgctt tgtgcatcta taataggtat ttgccatcta atggtgatga ttgggatgaa 2580

ccagctccag aagctccttg ggattctact ttgtttgctt gtcaaacttc aagagcttgg 2640

ttgaaaaatc ctggagagag aagatatgct gctgtttctg gtgtttctga aactaggtct 2700

tgttattctg ttttgttgtc tgaagctgaa ggtcattatg aaagagaaaa tagaatttct 2760

ttggatgaag aagctccaaa attgattgtt ttgagagctg attctcatga agaaattttg 2820

ggtaggttgg ataaaattag agaaagattt ttgcaaccaa ctggtgctgc tccaagagaa 2880

tctgaattga aagctcaagc tagaagaatt ttcttggaat tgttgggtga aactttggct 2940

caagatgctg cttcttctgg ttctcaaaaa ccattggctt tgtctttggt ttctactcca 3000

tctaaattgc aaagagaagt tgaattggct gctaaaggta ttccaagatg tttgaaaatg 3060

agaagagatt ggtcttctcc agctggttca agatatgctc cagaaccatt ggcttctgat 3120

agagttgctt tcatgtacgg tgaaggaagg tctccatact atggaatcac tcaagatatt 3180

catagaattt ggccagaatt gcatgaagtt attaacgaaa aaactaatag gttgtgggct 3240

gaaggtgata gatgggttat gccaagagct tcttttaaat ctgaattgga atctcaacaa 3300

caagaatttg atagaaatat gattgaaatg tttaggttgg gtattttgac ttctattgct 3360

tttactaatt tggctagaga tgttttgaat attactccaa aagctgcttt tggtttgtct 3420

ttgggtgaaa tttctatgat ttttgctttt tctaaaaaaa atggtttgat ttctgatcaa 3480

ttgactaaag atttgagaga atctgatgtt tggaacaaag cattggctgt tgaattcaat 3540

gctttgagag aagcatgggg tattccacaa tctgttccaa aagatgaatt ttggcaaggt 3600

tatattgtta gaggtactaa acaagatatt gaagctgcta ttgctccaga ttccaaatac 3660

gttaggttga ctatcattaa tgatgctaat actgctttga tttctggtaa accagatgct 3720

tgtaaagctg ctattgctag gttgggtggt aatattccag ctttgccagt tactcaagga 3780

atgtgtggtc attgtccaga agttggtcca tatactaaag atattgctaa aattcatgct 3840

aatttggaat ttccagttgt tgatggtttg gatttgtgga ctactattaa tcaaaaaaga 3900

ttggttccaa gagctactgg tgctaaagat gaatgggctc catcttcttt tggtgaatat 3960

gctggtcaac tttacgaaaa acaagctaat tttccacaaa ttgttgaaac tatctacaaa 4020

caaaattatg atgtttttgt tgaggttggt ccaaacaacc ataggtctac tgctgttaga 4080

actactttgg gtccacaaag aaatcatttg gctggtgcta ttgataaaca aaacgaagat 4140

gcttggacta ctattgttaa attggttgct tctttgaaag ctcatttggt tccaggtgtt 4200

actatttctc cattgtatca ttctaaattg gttgctgaag ctgaagcatg ttatgctgct 4260

ctgtgcaaag gagaaaaacc taagaagaac aaatttgtta gaaaaattca attgaatggt 4320

aggttcaatt ctaaagctga tccaatttct tctgctgatt tggcttcttt tccaccagct 4380

gatccagcta ttgaagctgc tatttcttca agaattatga aaccagttgc tccaaaattt 4440

tatgctaggt tgaatattga tgaacaagac gaaacaagag atccaatttt gaacaaagat 4500

aatgctccat ctagttcatc tagttcctct tcatctagtt cttcatctag ttctccatct 4560

ccagctcctt ctgctccagt tcaaaaaaaa gctgctccag ctgctgaaac taaagctgtt 4620

gcttctgctg atgctttgag atctgctttg ttggatttgg attctatgtt ggctttgtct 4680

tctgcttctg cttctggtaa tttggttgaa actgctccat ctgatgcttc tgttattgtt 4740

ccaccatgta atattgctga tttgggttca agagctttta tgaaaactta tggtgtttct 4800

gctccattgt acactggtgc tatggctaaa ggtattgctt ctgctgattt ggttattgct 4860

gctggtagac aaggcatttt ggcttctttt ggtgctggtg gtttgccaat gcaagttgtt 4920

agagaatcta ttgaaaaaat tcaagctgct ttgccaaatg gtccatatgc tgttaatttg 4980

attcattctc catttgattc taatttggaa aaaggtaatg ttgatttgtt tttggaaaaa 5040

ggtgttactt ttgttgaagc atctgctttt atgactttga ctccacaagt tgttaggtac 5100

agagctgctg gtttgactag aaatgctgat ggttctgtta atattagaaa tagaattatc 5160

ggaaaggttt caagaactga attggctgaa atgtttatga gacctgcccc agaacacttg 5220

ttgcaaaaat tgattgcttc tggtgaaatt aatcaagaac aagctgaatt ggctagaaga 5280

gttccagttg ctgatgatat tgctgttgaa gctgattctg gtggtcatac tgataataga 5340

ccaattcatg ttatcttgcc attgattatt aatttgagag acagattgca tagagaatgt 5400

ggttatccag ctaatttgag agttagagtt ggtgctggtg gtggtattgg ttgtccacaa 5460

gctgctttgg ctacttttaa tatgggtgct tctttcattg ttactggcac tgttaatcaa 5520

gttgctaaac aatctggtac ttgtgataat gttagaaaac aattggctaa agctacttat 5580

tctgatgttt gcatggctcc agctgctgat atgtttgaag aaggtgttaa attgcaagtt 5640

ttgaagaaag ggacaatgtt tccatcaaga gctaataagt tatacgaatt gttttgcaag 5700

tatgattctt ttgaatctat gccaccagct gaattggcta gagttgaaaa aagaattttc 5760

tcaagagctt tggaagaagt ttgggatgaa actaaaaatt tttacattaa taggttgcac 5820

aatccagaaa aaattcaaag agctgaaaga gatccaaaat tgaaaatgtc tttgtgtttt 5880

agatggtatt tgtctttggc ttcaagatgg gctaatactg gtgcttctga tagagttatg 5940

gattatcaag tttggtgtgg tccagctatt ggttctttta atgatttcat taaaggcacc 6000

tacttggacc cagctgttgc taacgaatat ccatgcgttg ttcaaattaa caaacaaatt 6060

ttgagaggtg cttgtttcct cagaagattg gaaattttga gaaatgctag gttgtctgat 6120

ggtgctgctg ctttggttgc ttctattgat gatacttatg ttccagctga aaaattgtaa 6180

<210> 37
<211> 6180
<212> DNA

<213> Artificial

<220>
<223> synthetic

<400> 37
atggccgctc gcaacgtgtc tgcagcgcat gagatgcacg atgaaaagcg catcgccgtc 60

gtcggcatgg ccgtccagta cgccggatgc aaaaccaagg acgagttctg ggaggtgctc 120

atgaacggca aggtcgagtc caaggtgatc agcgacaaac gactcggctc caactaccgc 180

gccgagcact acaaagcaga gcgcagcaag tatgccgaca ccttttgcaa cgaaacgtac 240

ggcacccttg acgagaacga gatcgacaac gagcacgaac tcctcctcaa cctcgccaag 300

caggcactcg cagagacatc cgtcaaagac tcgacacgct gcggcatcgt cagcggctgc 360

ctctcgttcc ccatggacaa cctccagggt gaactcctca acgtgtacca aaaccatgtc 420

gagaaaaagc tcggggcccg cgtcttcaag gacgcctccc attggtccga acgcgagcag 480

tccaacaaac ccgaggccgg tgaccgccgc atcttcatgg acccggcctc cttcgtcgcc 540

gaagaactca acctcggcgc ccttcactac tccgtcgacg cagcatgcgc cacggcgctc 600

tacgtgctcc gcctcgcgca ggatcatctc gtctccggcg ccgccgacgt catgctctgc 660

ggtgccacct gcctgccgga gccctttttc atcctttcgg gcttttccac cttccaggcc 720

atgcccgtcg gcacgggcca gaacgtgtcc atgccgctgc acaaggacag ccagggcctc 780

accccgggtg agggcggctc catcatggtc ctcaagcgtc tcgatgatgc catccgcgac 840

ggcgaccaca tctacggcac ccttctcggc gccaatgtca gcaactccgg cacaggtctg 900

cccctcaagc cccttctccc cagcgagaaa aagtgcctca tggacaccta cacgcgcatt 960

aacgtgcacc cgcacaagat tcagtacgtc gagtgccacg ccaccggcac gccccagggt 1020

gatcgtgtgg aaatcgacgc cgtcaaggcc tgctttgaag gcaaggtccc ccgtttcggt 1080

accacaaagg gcaactttgg acacaccctc gtcgcagccg gctttgccgg tatgtgcaag 1140

gtcctcctct ccatgaagca tggcatcatc ccgcccaccc cgggtatcga tgacgagacc 1200

aagatggacc ctctcgtcgt ctccggtgag gccatcccat ggccagagac caacggcgag 1260

cccaagcgcg ccggtctctc ggcctttggc tttggtggca ccaacgccca tgccgtcttt 1320

gaggagcatg acccctccaa cgccgcctgc acgggccacg actccatttc tgcgctctcg 1380

gcccgctgcg gcggtgaaag caacatgcgc atcgccatca ctggtatgga cgccaccttt 1440

ggcgctctca agggactcga cgccttcgag cgcgccattt acaccggcgc tcacggtgcc 1500

atcccactcc cagaaaagcg ctggcgcttt ctcggcaagg acaaggactt tcttgacctc 1560

tgcggcgtca aggccacccc gcacggctgc tacattgaag atgttgaggt cgacttccag 1620

cgcctccgca cgcccatgac ccctgaagac atgctcctcc ctcagcagct tctggccgtc 1680

accaccattg accgcgccat cctcgactcg ggaatgaaaa agggtggcaa tgtcgccgtc 1740

tttgtcggcc tcggcaccga cctcgagctc taccgtcacc gtgctcgcgt cgctctcaag 1800

gagcgcgtcc gccctgaagc ctccaagaag ctcaatgaca tgatgcagta cattaacgac 1860

tgcggcacat ccacatcgta cacctcgtac attggcaacc tcgtcgccac gcgcgtctcg 1920

tcgcagtggg gcttcacggg cccctccttt acgatcaccg agggcaacaa ctccgtctac 1980

cgctgcgccg agctcggcaa gtacctcctc gagaccggcg aggtcgatgg cgtcgtcgtt 2040

gcgggtgtcg atctctgcgg cagtgccgaa aacctttacg tcaagtctcg ccgcttcaag 2100

gtgtccacct ccgatacccc gcgcgccagc tttgacgccg ccgccgatgg ctactttgtc 2160

ggcgagggct gcggtgcctt tgtgctcaag cgtgagacta gctgcaccaa ggacgaccgt 2220

atctacgctt gcatggatgc catcgtccct ggcaacgtcc ctagcgcctg cttgcgcgag 2280

gccctcgacc aggcgcgcgt caagccgggc gatatcgaga tgctcgagct cagcgccgac 2340

tccgcccgcc acctcaagga cccgtccgtc ctgcccaagg agctcactgc cgaggaggaa 2400

atcggcggcc ttcagacgat ccttcgtgac gatgacaagc tcccgcgcaa cgtcgcaacg 2460

ggcagtgtca aggccaccgt cggtgacacc ggttatgcct ctggtgctgc cagcctcatc 2520

aaggctgcgc tttgcatcta caaccgctac ctgcccagca acggcgacga ctgggatgaa 2580

cccgcccctg aggcgccctg ggacagcacc ctctttgcgt gccagacctc gcgcgcttgg 2640

ctcaagaacc ctggcgagcg tcgctatgcg gccgtctcgg gcgtctccga gacgcgctcg 2700

tgctattccg tgctcctctc cgaagccgag ggccactacg agcgcgagaa ccgcatctcg 2760

ctcgacgagg aggcgcccaa gctcattgtg cttcgcgccg actcccacga ggagatcctt 2820

ggtcgcctcg acaagatccg cgagcgcttc ttgcagccca cgggcgccgc cccgcgcgag 2880

tccgagctca aggcgcaggc ccgccgcatc ttcctcgagc tcctcggcga gaccccttgcc 2940

caggatgccg cttcttcagg ctcgcaaaag cccctcgctc tcagcctcgt ctccacgccc 3000

tccaagctcc agcgcgaggt cgagctcgcg gccaagggta tcccgcgctg cctcaagatg 3060

cgccgcgatt ggagctcccc tgctggcagc cgctacgcgc ctgagccgct cgccagcgac 3120

cgcgtcgcct tcatgtacgg cgaaggtcgc agcccttact acggcatcac ccaagacatt 3180

caccgcattt ggcccgaact ccacgaggtc atcaacgaaa agacgaaccg tctctgggcc 3240

gaaggcgacc gctgggtcat gccgcgcgcc agcttcaagt cggagctcga gagccagcag 3300

caagagtttg atcgcaacat gattgaaatg ttccgtcttg gaatcctcac ctcaattgcc 3360

ttcaccaatc tggcgcgcga cgttctcaac atcacgccca aggccgcctt tggcctcagt 3420

cttggcgaga tttccatgat ttttgccttt tccaagaaga acggtctcat ctccgaccag 3480

ctcaccaagg atcttcgcga gtccgacgtg tggaacaagg ctctggccgt tgaatttaat 3540

gcgctgcgcg aggcctgggg cattccacag agtgtcccca aggacgagtt ctggcaaggc 3600

tacattgtgc gcggcaccaa gcaggatatc gaggcggcca tcgccccgga cagcaagtac 3660

gtgcgcctca ccatcatcaa tgatgccaac accgccctca ttagcggcaa gcccgacgcc 3720

tgcaaggctg cgatcgcgcg tctcggtggc aacattcctg cgcttcccgt gacccagggc 3780

atgtgcggcc actgccccga ggtgggacct tataccaagg atatcgccaa gatccatgcc 3840

aaccttgagt tccccgttgt cgacggcctt gacctctgga ccacaatcaa ccagaagcgc 3900

ctcgtgccac gcgccacggg cgccaaggac gaatgggccc cttcttcctt tggcgagtac 3960

gccggccagc tctacgagaa gcaggctaac ttcccccaaa tcgtcgagac catttacaag 4020

caaaactacg acgtctttgt cgaggttggg cccaacaacc accgtagcac cgcagtgcgc 4080

accacgcttg gtccccagcg caaccacctt gctggcgcca tcgacaagca gaacgaggat 4140

gcttggacga ccatcgtcaa gcttgtggct tcgctcaagg cccaccttgt tcctggcgtc 4200

acgatctcgc cgctgtacca ctccaagctt gtggcggagg ctgaggcttg ctacgctgcg 4260

ctctgcaagg gtgaaaagcc caagaagaac aagtttgtgc gcaagattca gctcaacggt 4320

cgcttcaaca gcaaggcgga ccccatctcc tcggccgatc ttgccagctt tccgcctgcg 4380

gaccctgcca ttgaagccgc catctcgagc cgcatcatga agccggttgc tccgaagttc 4440

tacgcgcgtc tcaacattga cgagcaggac gagacccgtg atccgatcct caacaaggac 4500

aacgcgccgt cttccagctc tagctcctct tccagctctt ccagctcttc cagcccgtcg 4560

ccagctccgt ccgccccagt gcaaaagaag gctgctccgg ccgcggagac caaggctgtt 4620

gcttcggctg acgcacttcg cagtgccctg ctcgatctcg acagtatgct tgcgctgagc 4680

tctgccagtg cctccggcaa ccttgttgag actgcgccta gcgacgcctc ggtcattgtg 4740

ccgccctgca acattgcgga tctcggcagc cgcgccttca tgaaaacgta cggtgtttcg 4800

gcgcctctgt acacgggcgc catggccaag ggcattgcct ctgcggacct cgtcattgcc 4860

gccggccgcc agggcatcct tgcgtccttt ggcgccggcg gacttcccat gcaggttgtg 4920

cgtgagtcca tcgaaaagat tcaggccgcc ctgcccaatg gcccgtacgc tgtcaacctt 4980

atccattctc cctttgacag caacctcgaa aagggcaatg tcgatctctt cctcgagaag 5040

ggtgtcacct ttgtcgaggc ctcggccttt atgacgctca ccccgcaggt cgtgcggtac 5100

cgcgcggctg gcctcacgcg caacgccgac ggctcggtca acatccgcaa ccgtatcatt 5160

ggcaaggtct cgcgcaccga gctcgccgag atgttcatgc gtcctgcgcc cgagcacctt 5220

cttcagaagc tcattgcttc cggcgagatc aaccaggagc aggccgagct cgcccgccgt 5280

gttcccgtcg ctgacgacat cgcggtcgaa gctgactcgg gtggccacac cgacaaccgc 5340

cccatccacg tcattctgcc cctcatcatc aaccttcgcg accgccttca ccgcgagtgc 5400

ggctacccgg ccaaccttcg cgtccgtgtg ggcgccggcg gtggcattgg gtgcccccag 5460

gcggcgctgg ccaccttcaa catgggtgcc tcctttattg tcaccggcac cgtgaaccag 5520

gtcgccaagc agtcgggcac gtgcgacaat gtgcgcaagc agctcgcgaa ggccacttac 5580

tcggacgtat gcatggcccc ggctgccgac atgttcgagg aaggcgtcaa gcttcaggtc 5640

ctcaagaagg gaaccatgtt tccctcgcgc gccaacaagc tctacgagct cttttgcaag 5700

tacgactcgt tcgagtccat gccccccgca gagcttgcgc gcgtcgagaa gcgcatcttc 5760

agccgcgcgc tcgaagaggt ctgggacgag accaaaaact tttacattaa ccgtcttcac 5820

aacccggaga agatccagcg cgccgagcgc gaccccaagc tcaagatgtc gctgtgcttt 5880

cgctggtacc tgagcctggc gagccgctgg gccaacactg gagcttccga tcgcgtcatg 5940

gactaccagg tctggtgcgg tcctgccatt ggttccttca acgatttcat caagggaact 6000

taccttgatc cggccgtcgc aaacgagtac ccgtgcgtcg ttcagattaa caagcagatc 6060

cttcgtggag cgtgcttctt gcgccgtctc gaaattctgc gcaacgcacg cctttccgat 6120

ggcgctgccg ctcttgtggc cagcatcgat gacacatacg tcccggccga gaagctgtaa 6180

<210> 38
<211> 8436
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(8433)

<400> 38

```
atg aag gac atg gaa gat aga cgg gtc gct att gtg ggc atg tca gct       48
Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1               5                  10                  15

cac ttg cct tgt ggg aca gat gtg aag gaa tca tgg cag gct att cgc       96
His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
            20                  25                  30

gat gga atc gac tgt cta agt gac cta ccc gcg gat cgt ctc gac gtt      144
Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
        35                  40                  45

aca gct tac tac aat ccc aac aaa gcc acg aaa gac aag atc tac tgc      192
Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
    50                  55                  60

aaa cgg ggt ggc ttc atc ccg aac tat gac ttc gac ccc cgc gaa ttt      240
Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65                  70                  75                  80

ggg ctc aac atg ttt caa atg gaa gac tct gat gcg aat cag aca ctt      288
Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
                85                  90                  95

acc ttg ctc aaa gtc aaa caa gct ctc gaa gat gca agc ata gag cct      336
Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro
            100                 105                 110

ttc acc aag gag aag aag aac att gga tgt gtt tta ggt att ggt ggg      384
Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
        115                 120                 125

ggc caa aag gcg agt cat gag ttc tac tct cgt ctc aac tac gtt gtc      432
Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
    130                 135                 140

gtt gaa aag gta ctt cgg aaa atg ggt tta cca gat gct gat gtt gaa      480
Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145                 150                 155                 160

gaa gct gtg gag aaa tac aag gca aat ttt ccc gag tgg cgc cta gac      528
Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
                165                 170                 175

tct ttc cct ggg ttt ctt ggg aat gta acg gct ggt cgg tgc agt aac      576
Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
            180                 185                 190
```

```
acc ttc aac atg gaa ggt atg aac tgc gtt gtg gat gct gca tgt gcc      624
Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala
        195                 200                 205

agt tct cta att gca atc aag gtt gca gtt gaa gag cta ctc ttt ggt      672
Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
    210                 215                 220

gac tgt gac acc atg att gca ggt gcc acc tgc acg gac aat tca ctt      720
Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
225                 230                 235                 240

ggc atg tac atg gcc ttc tct aaa acg cca gtt ttt tct act gac cca      768
Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
                245                 250                 255

agt gtc cgc gcg tat gat gag aaa aca aaa ggg atg cta att gga gaa      816
Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
            260                 265                 270

ggt tca gca atg ttc gtt ctt aaa cgc tat gcg gat gcc gta cgt gat      864
Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
        275                 280                 285

ggc gac aca att cac gcg gtt ctg cgt tct tgc tct tcg tct agt gat      912
Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp
    290                 295                 300

gga aaa gcg gca gga att tat act cct act ata tct gga caa gaa gaa      960
Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
305                 310                 315                 320

gct ttg cgt cga gcg tat gcc cgt gcg ggg gta tgt cca tct acg atc     1008
Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
                325                 330                 335

ggg ctt gtt gag ggt cac ggg aca ggg acc cct gtt gga gat cgc att     1056
Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
            340                 345                 350

gag tta aca gct ctg cgg aac ttg ttt gac aaa gct ttt ggt agc aag     1104
Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
        355                 360                 365

aag gaa caa ata gca gtt ggc agc ata aag tct cag ata ggt cac ctg     1152
Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
    370                 375                 380

aaa tct gtt gcc ggc ttt gcc ggc ttg gtc aaa gct gtg ctt gcg ctt     1200
Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385                 390                 395                 400

aaa cac aaa acg ctc cca ggt tcg att aat gtc gac cag cca cct ttg     1248
Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
                405                 410                 415

ttg tat gac ggt act caa att caa gac tct tct tta tat atc aac aag     1296
Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
            420                 425                 430

aca aat aga cca tgg ttt acg caa aac aag ctt ccg cgt cgg gct ggt     1344
Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
        435                 440                 445

gtc tca agt ttt gga ttt gga ggt gca aac tac cac gcg gtt ctg gaa     1392
Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
    450                 455                 460
```

```
gaa ttc gag ccc gag cat gaa aaa cca tac cgc ctc aat act gtt gga     1440
Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465                 470                 475                 480

cat cct gtc ctc ttg tac gct ccg tct gtg gaa gcc ctc aaa gta ctt     1488
His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
                485                 490                 495

tgc aac gac cag ctt gcg gag ctc aca att gca ttg gaa gag gca aaa     1536
Cys Asn Asp Gln Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys
            500                 505                 510

aca cat aaa aat gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa     1584
Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu
        515                 520                 525

ttt cag ctc caa gga agc tgt cct cca gaa aat ccg aga gta gga ttt     1632
Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe
    530                 535                 540

tta gca aca ctg cct act tca aat atc att gtc gcg ctt aag gca att     1680
Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile
545                 550                 555                 560

ctc gcg cag ctt gat gca aaa cca gat gcg aag aaa tgg gat ttg cct     1728
Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro
                565                 570                 575

cat aaa aag gct ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc     1776
His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly
            580                 585                 590

tct gtt gct gcg ctc ttc gca gga cag ggt acc cag tac tta aac atg     1824
Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met
        595                 600                 605

ttc tct gat gtg gca atg aac tgg cca ccg ttc cgt gac agc att gtc     1872
Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val
    610                 615                 620

gca atg gaa gaa gct caa act gag gta ttt gag ggc caa gtt gaa cca     1920
Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro
625                 630                 635                 640

att agc aaa gtt ctg ttt cca cga gag cgc tat gca tcc gaa agt gaa     1968
Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu
                645                 650                 655

cag ggg aat gaa ctt ctt tgc tta aca gag tac tct cag cca act acg     2016
Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr
            660                 665                 670

ata gca gcc gca gta ggg gcc ttc gat att ttc aaa gcg gct ggc ttt     2064
Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe
        675                 680                 685

aag cca gac atg gtt gga ggg cat tca ctt ggc gaa ttt gct gct ttg     2112
Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu
    690                 695                 700

tac gcg gct ggg tcc att tcg cgt gac gac ctg tac aag ctt gtg tgc     2160
Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys
705                 710                 715                 720

aaa cgg gca aag gca atg gcg aac gct agt gac gga gct atg gca gca     2208
Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala
                725                 730                 735
```

```
gtg att ggc cca gat gca cgt cta gtt acg cca caa aat agt gac gtt     2256
Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val
            740                 745                 750

tat gtc gca aac ttc aac tcc gca act caa gta gtc atc agt ggc act     2304
Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr
        755                 760                 765

gtt caa ggt gtg aaa gaa gag tcg aaa ttg ctc att tca aag ggg ttc     2352
Val Gln Gly Val Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe
    770                 775                 780

cgc gta ctg cca ctt aaa tgc cag ggc gcc ttc cat tct cct ttg atg     2400
Arg Val Leu Pro Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met
785                 790                 795                 800

ggg cct tct gag gat agt ttc aaa tca ctt gtg gag act tgt acc atc     2448
Gly Pro Ser Glu Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile
                805                 810                 815

tcg ccg cca aaa aat gtg aaa ttc ttt tgc aat gtt agt ggc aag gaa     2496
Ser Pro Pro Lys Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu
            820                 825                 830

agc cca aac cca aaa cag acc ctc aag tca cac atg acg tct agc gtt     2544
Ser Pro Asn Pro Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val
        835                 840                 845

cag ttc gag gag cag att cgt aac atg tac gat gcc gga gca cgt gtt     2592
Gln Phe Glu Glu Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val
    850                 855                 860

ttt ctg gag ttt gga ccc cgc caa gtc ctt gca aag ctt atc gcg gaa     2640
Phe Leu Glu Phe Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu
865                 870                 875                 880

atg ttt ccc tcg tgt aca gct atc agc gtt aac ccc gcg agc agt ggt     2688
Met Phe Pro Ser Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly
                885                 890                 895

gac agt gac gtg caa ctc cgc ctc gcc gcc gta aaa ttc gcg gtc tcg     2736
Asp Ser Asp Val Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser
            900                 905                 910

ggt gca gcc ctt agc acc ttt gat cca tgg gag tat cgc aag cca caa     2784
Gly Ala Ala Leu Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln
        915                 920                 925

gat ctt ctt att cga aaa cca cga aaa act gcc ctt gtt cta tca gca     2832
Asp Leu Leu Ile Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala
    930                 935                 940

gca aca tat gtt tcc cca aag act ctt gca gaa cgt aaa aag gct atg     2880
Ala Thr Tyr Val Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met
945                 950                 955                 960

gaa gat atc aag cta gta tcc att aca cca aga gat agt atg gta tca     2928
Glu Asp Ile Lys Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser
                965                 970                 975

att gga aaa atc gcg caa gaa gta cgg aca gct aaa cag cct tta gaa     2976
Ile Gly Lys Ile Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu
            980                 985                 990

acc gaa att cga aga ctc aac aaa  gaa tta gaa cat ctc  aag aga gag   3024
Thr Glu Ile Arg Arg Leu Asn Lys  Glu Leu Glu His Leu  Lys Arg Glu
        995                 1000                1005
```

```
cta gca  gca gcc aaa gcg agt  gtc aag tct gca tca  aaa agc tct      3069
Leu Ala  Ala Ala Lys Ala Ser  Val Lys Ser Ala Ser  Lys Ser Ser
    1010                 1015                 1020

aaa gag  cga tct gtc cta tca  aag cac cgc gct ttg  ctt caa aac      3114
Lys Glu  Arg Ser Val Leu Ser  Lys His Arg Ala Leu  Leu Gln Asn
    1025                 1030                 1035

att ttg  caa gac tac gat gat  ctt cgt gtg gtg cca  ttc gct gtt      3159
Ile Leu  Gln Asp Tyr Asp Asp  Leu Arg Val Val Pro  Phe Ala Val
    1040                 1045                 1050

cgt tct  gtt gca gtg gac aac  acc gcg ccg tat gct  gac caa gtt      3204
Arg Ser  Val Ala Val Asp Asn  Thr Ala Pro Tyr Ala  Asp Gln Val
    1055                 1060                 1065

tcg acc  cca gcg tca gag cgg  tcg gct tca ccg ctt  ttc gag aaa      3249
Ser Thr  Pro Ala Ser Glu Arg  Ser Ala Ser Pro Leu  Phe Glu Lys
    1070                 1075                 1080

cgc agt  tcg gtt tcg tca gca  cgc ctc gct gaa gct  gaa gcc gcg      3294
Arg Ser  Ser Val Ser Ser Ala  Arg Leu Ala Glu Ala  Glu Ala Ala
    1085                 1090                 1095

gta ctg  agc gtt ctc gca gac  aag aca ggc tac gac  agc tca atg      3339
Val Leu  Ser Val Leu Ala Asp  Lys Thr Gly Tyr Asp  Ser Ser Met
    1100                 1105                 1110

atc gag  atg gac atg gac ctg  gag agt gag ctt ggc  gtt gat agc      3384
Ile Glu  Met Asp Met Asp Leu  Glu Ser Glu Leu Gly  Val Asp Ser
    1115                 1120                 1125

atc aaa  cgc gtg gag atc atg  agc gag gtt caa acg  ctg ctc agc      3429
Ile Lys  Arg Val Glu Ile Met  Ser Glu Val Gln Thr  Leu Leu Ser
    1130                 1135                 1140

gtg gaa  gtc tcc gac gtt gac  gct ctg tca aga acc  aag act gtt      3474
Val Glu  Val Ser Asp Val Asp  Ala Leu Ser Arg Thr  Lys Thr Val
    1145                 1150                 1155

ggc gac  gtc atc gag gcg atg  aag ctg gaa ctc ggt  gga ccc caa      3519
Gly Asp  Val Ile Glu Ala Met  Lys Leu Glu Leu Gly  Gly Pro Gln
    1160                 1165                 1170

ggc cag  act ttg acc gcg gaa  tcg atc cgt cag cca  ccg gtg tcc      3564
Gly Gln  Thr Leu Thr Ala Glu  Ser Ile Arg Gln Pro  Pro Val Ser
    1175                 1180                 1185

gag cct  gct gta ccg acc tca  tcg tca agc agt att  gct aat gtt      3609
Glu Pro  Ala Val Pro Thr Ser  Ser Ser Ser Ser Ile  Ala Asn Val
    1190                 1195                 1200

tcg tca  gca cgc ctc gct gaa  gct gaa gct gcg gta  ctg agc gtt      3654
Ser Ser  Ala Arg Leu Ala Glu  Ala Glu Ala Ala Val  Leu Ser Val
    1205                 1210                 1215

ctc gca  gac aag aca ggc tac  gac agc tca atg atc  gag atg gac      3699
Leu Ala  Asp Lys Thr Gly Tyr  Asp Ser Ser Met Ile  Glu Met Asp
    1220                 1225                 1230

atg gac  ctg gag agc gag ctt  ggc gtt gat agc atc  aaa cgc gtg      3744
Met Asp  Leu Glu Ser Glu Leu  Gly Val Asp Ser Ile  Lys Arg Val
    1235                 1240                 1245

gag atc  atg agc gag gtt caa  acg ctg ctc agc gtg  gaa gtc tcc      3789
Glu Ile  Met Ser Glu Val Gln  Thr Leu Leu Ser Val  Glu Val Ser
    1250                 1255                 1260
```

```
gac gtt  gac gct ctg tca aga  act aag act gtt ggc  gac gtc atc      3834
Asp Val  Asp Ala Leu Ser Arg  Thr Lys Thr Val Gly  Asp Val Ile
    1265                 1270                 1275

gag gcg  atg aag ctg gaa ctc  ggt gga ccc caa ggc  cag act ttg      3879
Glu Ala  Met Lys Leu Glu Leu  Gly Gly Pro Gln Gly  Gln Thr Leu
    1280                 1285                 1290

acc gcg  gaa tcg atc cgt cag  cca ccg gtg tct gag  cct gct gta      3924
Thr Ala  Glu Ser Ile Arg Gln  Pro Pro Val Ser Glu  Pro Ala Val
    1295                 1300                 1305

ccg acc  tca tcg tca agc agt  att gct aat gtt tcg  tca gca cgc      3969
Pro Thr  Ser Ser Ser Ser Ser  Ile Ala Asn Val Ser  Ser Ala Arg
    1310                 1315                 1320

ctc gct  gaa gct gaa gcg gcg  gta ctg agc gtt ctc  gca gac aag      4014
Leu Ala  Glu Ala Glu Ala Ala  Val Leu Ser Val Leu  Ala Asp Lys
    1325                 1330                 1335

aca ggc  tac gac agc tca atg  atc gag atg gac atg  gac ctg gag      4059
Thr Gly  Tyr Asp Ser Ser Met  Ile Glu Met Asp Met  Asp Leu Glu
    1340                 1345                 1350

agc gag  ctt ggc gtc gac agc  atc aaa cgc gtg gag  atc atg agc      4104
Ser Glu  Leu Gly Val Asp Ser  Ile Lys Arg Val Glu  Ile Met Ser
    1355                 1360                 1365

gag gtt  caa acg ctg ctc agc  gtg gaa gtc tcc gac  gtt gac gct      4149
Glu Val  Gln Thr Leu Leu Ser  Val Glu Val Ser Asp  Val Asp Ala
    1370                 1375                 1380

ctg tca  aga acc aag act gtt  ggc gac gtc atc gag  gcg atg aag      4194
Leu Ser  Arg Thr Lys Thr Val  Gly Asp Val Ile Glu  Ala Met Lys
    1385                 1390                 1395

ctg gaa  ctc ggt gga ccc caa  ggc cag act ttg acc  gcg gaa tcg      4239
Leu Glu  Leu Gly Gly Pro Gln  Gly Gln Thr Leu Thr  Ala Glu Ser
    1400                 1405                 1410

atc cgt  cag cca ccg gtg tcc  gag cct gct gta ccg  acc tca tcg      4284
Ile Arg  Gln Pro Pro Val Ser  Glu Pro Ala Val Pro  Thr Ser Ser
    1415                 1420                 1425

tca agc  agt att gct aat gtt  ttg tca gca cgc ctc  gct gaa gct      4329
Ser Ser  Ser Ile Ala Asn Val  Leu Ser Ala Arg Leu  Ala Glu Ala
    1430                 1435                 1440

gaa gcc  gcg gta ctg agc gtt  ctc gca gac aag aca  ggc tac gac      4374
Glu Ala  Ala Val Leu Ser Val  Leu Ala Asp Lys Thr  Gly Tyr Asp
    1445                 1450                 1455

agc tca  atg atc gag atg gac  atg gac ctg gag agc  gag ctt ggc      4419
Ser Ser  Met Ile Glu Met Asp  Met Asp Leu Glu Ser  Glu Leu Gly
    1460                 1465                 1470

gtt gat  agc atc aaa cgc gtg  gag atc atg agc gag  gtt caa acg      4464
Val Asp  Ser Ile Lys Arg Val  Glu Ile Met Ser Glu  Val Gln Thr
    1475                 1480                 1485

ttg ctc  agc gtg gaa gtc tcc  gac gtt gac gct ctg  tca aga acc      4509
Leu Leu  Ser Val Glu Val Ser  Asp Val Asp Ala Leu  Ser Arg Thr
    1490                 1495                 1500

aag act  gtt ggc gac gtc atc  gag gcg atg aag ctg  gaa ctc ggt      4554
Lys Thr  Val Gly Asp Val Ile  Glu Ala Met Lys Leu  Glu Leu Gly
    1505                 1510                 1515
```

```
gga ccc  caa ggc cag act ttg  acc gcg gaa tcg atc  cgt cag cca      4599
Gly Pro  Gln Gly Gln Thr Leu  Thr Ala Glu Ser Ile  Arg Gln Pro
    1520                 1525                 1530

ccg gtg  tct gag cct gct gta  ccg acc tca tcg tca  agc agt att      4644
Pro Val  Ser Glu Pro Ala Val  Pro Thr Ser Ser Ser  Ser Ser Ile
    1535                 1540                 1545

gct aat  gtt tcg tca gca cgc  ctc gct gaa gct gaa  gcc gcg gta      4689
Ala Asn  Val Ser Ser Ala Arg  Leu Ala Glu Ala Glu  Ala Ala Val
    1550                 1555                 1560

ctg agc  gtt ctc gca gac aag  aca ggc tac gac agc  tca atg atc      4734
Leu Ser  Val Leu Ala Asp Lys  Thr Gly Tyr Asp Ser  Ser Met Ile
    1565                 1570                 1575

gag atg  gac atg gac ctg gag  agt gag ctt ggc gtc  gac agc atc      4779
Glu Met  Asp Met Asp Leu Glu  Ser Glu Leu Gly Val  Asp Ser Ile
    1580                 1585                 1590

aaa cgc  gtg gag atc atg agc  gag gtt caa acg ctg  ctc agc gtg      4824
Lys Arg  Val Glu Ile Met Ser  Glu Val Gln Thr Leu  Leu Ser Val
    1595                 1600                 1605

gaa gtc  tcc gac gtt gac gct  ctg tca aga acc aag  act gtt ggc      4869
Glu Val  Ser Asp Val Asp Ala  Leu Ser Arg Thr Lys  Thr Val Gly
    1610                 1615                 1620

gac gtc  atc gag gcg atg aag  ctg gaa ctc ggt gga  ccc caa ggc      4914
Asp Val  Ile Glu Ala Met Lys  Leu Glu Leu Gly Gly  Pro Gln Gly
    1625                 1630                 1635

cag act  ttg acc tct gaa ccg  atc cat cag cca cca  gtg tcc gag      4959
Gln Thr  Leu Thr Ser Glu Pro  Ile His Gln Pro Pro  Val Ser Glu
    1640                 1645                 1650

cct gct  gta ccg acc tca tcg  tca agc agt att gct  aat gtt tct      5004
Pro Ala  Val Pro Thr Ser Ser  Ser Ser Ser Ile Ala  Asn Val Ser
    1655                 1660                 1665

tca gca  cgc ctc gct gaa gct  gaa gcc gcg gta ctg  agc gtt ctc      5049
Ser Ala  Arg Leu Ala Glu Ala  Glu Ala Ala Val Leu  Ser Val Leu
    1670                 1675                 1680

gca gac  aag aca ggc tac gac  agc tca atg atc gag  atg gac atg      5094
Ala Asp  Lys Thr Gly Tyr Asp  Ser Ser Met Ile Glu  Met Asp Met
    1685                 1690                 1695

gac ctg  gag agc gag ctt ggc  gtt gat agc atc aaa  cgc gtg gaa      5139
Asp Leu  Glu Ser Glu Leu Gly  Val Asp Ser Ile Lys  Arg Val Glu
    1700                 1705                 1710

atc atg  agc gag gtt caa acg  ctg ctc agc gtg gaa  gtc tcc gac      5184
Ile Met  Ser Glu Val Gln Thr  Leu Leu Ser Val Glu  Val Ser Asp
    1715                 1720                 1725

gtt gac  gct ctg tca aga acc  aag act gtt ggc gac  gtc atc gag      5229
Val Asp  Ala Leu Ser Arg Thr  Lys Thr Val Gly Asp  Val Ile Glu
    1730                 1735                 1740

gcg atg  aag atg gaa ctc ggt  gga ccc caa ggc cag  act ttg acc      5274
Ala Met  Lys Met Glu Leu Gly  Gly Pro Gln Gly Gln  Thr Leu Thr
    1745                 1750                 1755

gcg gaa  tcg atc cgt cag cca  ccg gtg tct gag cct  gct gta ccg      5319
Ala Glu  Ser Ile Arg Gln Pro  Pro Val Ser Glu Pro  Ala Val Pro
    1760                 1765                 1770
```

```
acc tca  tcg tca agc agt att  gct aat gtt tcg tca  gca cgc ctc      5364
Thr Ser  Ser Ser Ser Ser Ile  Ala Asn Val Ser Ser  Ala Arg Leu
    1775                 1780                 1785

gct gaa  gct gaa gcg gcg gta  ctg agc gtt ctc gca  gac aag aca      5409
Ala Glu  Ala Glu Ala Ala Val  Leu Ser Val Leu Ala  Asp Lys Thr
    1790                 1795                 1800

ggc tac  gac agc tca atg atc  gag atg gac atg gac  ctg gag agc      5454
Gly Tyr  Asp Ser Ser Met Ile  Glu Met Asp Met Asp  Leu Glu Ser
    1805                 1810                 1815

gag ctt  ggc gtt gat agc atc  aaa cgc gtg gag atc  atg agc gag      5499
Glu Leu  Gly Val Asp Ser Ile  Lys Arg Val Glu Ile  Met Ser Glu
    1820                 1825                 1830

gtt caa  gcg ctg ctc agc gtg  gaa gtc tcc gac gtt  gac gct ctg      5544
Val Gln  Ala Leu Leu Ser Val  Glu Val Ser Asp Val  Asp Ala Leu
    1835                 1840                 1845

tca aga  acc aag act gtt ggc  gac gtc atc gag gcg  atg aag atg      5589
Ser Arg  Thr Lys Thr Val Gly  Asp Val Ile Glu Ala  Met Lys Met
    1850                 1855                 1860

gaa ctc  ggt gga ccc caa ggc  cag act ttg acc gca  gaa tcg atc      5634
Glu Leu  Gly Gly Pro Gln Gly  Gln Thr Leu Thr Ala  Glu Ser Ile
    1865                 1870                 1875

cgt gag  cca ccg gtg tct gag  cct gct gta ccg acc  tca tcg tca      5679
Arg Glu  Pro Pro Val Ser Glu  Pro Ala Val Pro Thr  Ser Ser Ser
    1880                 1885                 1890

agt agt  atc gct aat gtt tct  tca gct cgc ctc gct  gaa gct gaa      5724
Ser Ser  Ile Ala Asn Val Ser  Ser Ala Arg Leu Ala  Glu Ala Glu
    1895                 1900                 1905

gcc gcg  gta ctg agc gtt ctc  gca gac aag aca ggc  tac gac agc      5769
Ala Ala  Val Leu Ser Val Leu  Ala Asp Lys Thr Gly  Tyr Asp Ser
    1910                 1915                 1920

tca atg  atc gag atg gac atg  gac ctg gag agt gag  ctt ggc gtc      5814
Ser Met  Ile Glu Met Asp Met  Asp Leu Glu Ser Glu  Leu Gly Val
    1925                 1930                 1935

gac agc  atc aaa cgc gtg gag  atc atg agc gag gtt  caa acg ttg      5859
Asp Ser  Ile Lys Arg Val Glu  Ile Met Ser Glu Val  Gln Thr Leu
    1940                 1945                 1950

ctc agc  gtg gaa gtc tcc gac  gtt gac gct ctg tca  aga acc aag      5904
Leu Ser  Val Glu Val Ser Asp  Val Asp Ala Leu Ser  Arg Thr Lys
    1955                 1960                 1965

act gtt  ggc gac gtc atc gag  gcg atg aag ctg gaa  ctt ggg gaa      5949
Thr Val  Gly Asp Val Ile Glu  Ala Met Lys Leu Glu  Leu Gly Glu
    1970                 1975                 1980

tca tca  agt att gag act ctc  aat tgt acc gag gtt  gag cac acg      5994
Ser Ser  Ser Ile Glu Thr Leu  Asn Cys Thr Glu Val  Glu His Thr
    1985                 1990                 1995

agc tac  aaa agt gtc aag gct  tca ggg tgt gag aat  gta gat acc      6039
Ser Tyr  Lys Ser Val Lys Ala  Ser Gly Cys Glu Asn  Val Asp Thr
    2000                 2005                 2010

cgt ttc  gct aag gtt gta caa  atc tcg ctt cct agc  aag ctg aaa      6084
Arg Phe  Ala Lys Val Val Gln  Ile Ser Leu Pro Ser  Lys Leu Lys
    2015                 2020                 2025
```

```
tcc act  gtg tcg cac gat cga  cct gta att gtt gta  gat gat gga      6129
Ser Thr  Val Ser His Asp Arg  Pro Val Ile Val Val  Asp Asp Gly
    2030                 2035                 2040

acg ccc  tta acc acg gag ctt  tgt aaa att ctt ggg  ggt aat att      6174
Thr Pro  Leu Thr Thr Glu Leu  Cys Lys Ile Leu Gly  Gly Asn Ile
    2045                 2050                 2055

gtg gtt  ctc tct tat caa ggg  aag ccc gct ggt cca  cgg gga gtc      6219
Val Val  Leu Ser Tyr Gln Gly  Lys Pro Ala Gly Pro  Arg Gly Val
    2060                 2065                 2070

gag gtg  cca gat ctt tcc gag  gaa gcc cta att caa  gct ctt gca      6264
Glu Val  Pro Asp Leu Ser Glu  Glu Ala Leu Ile Gln  Ala Leu Ala
    2075                 2080                 2085

ttg att  cgg tct aca tat gga  gtt cca att ggt ttt  att tgt cag      6309
Leu Ile  Arg Ser Thr Tyr Gly  Val Pro Ile Gly Phe  Ile Cys Gln
    2090                 2095                 2100

caa gtg  tct aat gtg agc acc  aag gca cag ctt tgt  tgg gca ctc      6354
Gln Val  Ser Asn Val Ser Thr  Lys Ala Gln Leu Cys  Trp Ala Leu
    2105                 2110                 2115

ctc gca  gcg aag cat ctc aag  aag gat ttg aat gct  gtc tta ccc      6399
Leu Ala  Ala Lys His Leu Lys  Lys Asp Leu Asn Ala  Val Leu Pro
    2120                 2125                 2130

gat tca  aga tcc ttc ttc gtc  gga gtt gta cgc ttg  aac ggg aaa      6444
Asp Ser  Arg Ser Phe Phe Val  Gly Val Val Arg Leu  Asn Gly Lys
    2135                 2140                 2145

ctt gga  act ttc gaa aac atc  agc gac ttc tct aaa  ttt gat ttg      6489
Leu Gly  Thr Phe Glu Asn Ile  Ser Asp Phe Ser Lys  Phe Asp Leu
    2150                 2155                 2160

acg aaa  gcc cta gat tac gga  cag cgt ggt tct ctc  tta ggc ctg      6534
Thr Lys  Ala Leu Asp Tyr Gly  Gln Arg Gly Ser Leu  Leu Gly Leu
    2165                 2170                 2175

tgc aag  tca cta gac tta gaa  tgg gaa cag gtg ttt  tgc cgt gga      6579
Cys Lys  Ser Leu Asp Leu Glu  Trp Glu Gln Val Phe  Cys Arg Gly
    2180                 2185                 2190

ata gat  ctt gcg tgt gat ctt  atg cca ctc cag gcc  gca agg ata      6624
Ile Asp  Leu Ala Cys Asp Leu  Met Pro Leu Gln Ala  Ala Arg Ile
    2195                 2200                 2205

ctc aga  aat gag ctt cag tgt  ccc aat atg cgc ctt  cgc gag gtt      6669
Leu Arg  Asn Glu Leu Gln Cys  Pro Asn Met Arg Leu  Arg Glu Val
    2210                 2215                 2220

ggg tac  gat att tct ggc gcc  agg tac acc att tca  acc gat gac      6714
Gly Tyr  Asp Ile Ser Gly Ala  Arg Tyr Thr Ile Ser  Thr Asp Asp
    2225                 2230                 2235

ctg cta  tgt gga ccc tcg aag  gct aaa gta gag gcc  gca gac ttg      6759
Leu Leu  Cys Gly Pro Ser Lys  Ala Lys Val Glu Ala  Ala Asp Leu
    2240                 2245                 2250

ttt ctt  gtg aca ggt ggc gca  cga ggt att aca cct  cat tgt gtt      6804
Phe Leu  Val Thr Gly Gly Ala  Arg Gly Ile Thr Pro  His Cys Val
    2255                 2260                 2265

cgt gag  att gca agt cga tcc  ccc gga acc aca ttt  gtg ctg gtt      6849
Arg Glu  Ile Ala Ser Arg Ser  Pro Gly Thr Thr Phe  Val Leu Val
    2270                 2275                 2280
```

```
gga aga  agc gaa atg tcc gac  gag cct gac tgg gct  gtt ggc cac      6894
Gly Arg  Ser Glu Met Ser Asp  Glu Pro Asp Trp Ala  Val Gly His
    2285                 2290                 2295

tac aat  aaa gac ctg gac caa  agc aca atg aaa cac  ttg aaa gca      6939
Tyr Asn  Lys Asp Leu Asp Gln  Ser Thr Met Lys His  Leu Lys Ala
    2300                 2305                 2310

acg cat  gct gct gga ggg gta  aaa cct acg cct aaa  gca cat cgt      6984
Thr His  Ala Ala Gly Gly Val  Lys Pro Thr Pro Lys  Ala His Arg
    2315                 2320                 2325

gca ctt  gtg aac agg gtc act  ggc tca cgg gag gta  cga gaa tct      7029
Ala Leu  Val Asn Arg Val Thr  Gly Ser Arg Glu Val  Arg Glu Ser
    2330                 2335                 2340

ctt aga  gca atc cag gag gca  ggg gca aat gtc gaa  tat atc gcc      7074
Leu Arg  Ala Ile Gln Glu Ala  Gly Ala Asn Val Glu  Tyr Ile Ala
    2345                 2350                 2355

tgt gat  gtt tcg gat gaa aac  aag gtc cgc caa ctt  gtg caa aga      7119
Cys Asp  Val Ser Asp Glu Asn  Lys Val Arg Gln Leu  Val Gln Arg
    2360                 2365                 2370

gtg gag  caa aag tat ggc tgt  gaa ata act ggg att  tgg cat gca      7164
Val Glu  Gln Lys Tyr Gly Cys  Glu Ile Thr Gly Ile  Trp His Ala
    2375                 2380                 2385

agc ggg  gtt ctt cgt gac aaa  ctt gtc gag caa aag  act aca gac      7209
Ser Gly  Val Leu Arg Asp Lys  Leu Val Glu Gln Lys  Thr Thr Asp
    2390                 2395                 2400

gac ttt  gag gca gtt ttt ggg  acc aag gtg act ggc  ctt gta aac      7254
Asp Phe  Glu Ala Val Phe Gly  Thr Lys Val Thr Gly  Leu Val Asn
    2405                 2410                 2415

atc gtg  tca caa gtc aat atg  tct aag cta cga cac  ttc atc ctc      7299
Ile Val  Ser Gln Val Asn Met  Ser Lys Leu Arg His  Phe Ile Leu
    2420                 2425                 2430

ttc agt  tct ttg gct gga ttt  cat ggg aac aag ggc  caa acg gat      7344
Phe Ser  Ser Leu Ala Gly Phe  His Gly Asn Lys Gly  Gln Thr Asp
    2435                 2440                 2445

tat gca  att gct aat gaa gcc  ttg aac aaa atc gcg  cat act ctc      7389
Tyr Ala  Ile Ala Asn Glu Ala  Leu Asn Lys Ile Ala  His Thr Leu
    2450                 2455                 2460

tca gcg  ttt ttg ccc aaa ctg  aat gca aag gtg cta  gac ttc ggt      7434
Ser Ala  Phe Leu Pro Lys Leu  Asn Ala Lys Val Leu  Asp Phe Gly
    2465                 2470                 2475

ccg tgg  gta ggt tca gga atg  gta acc gaa aca ctt  gag aag cat      7479
Pro Trp  Val Gly Ser Gly Met  Val Thr Glu Thr Leu  Glu Lys His
    2480                 2485                 2490

ttt aaa  gct atg ggg gtt cag  act att cct ctc gag  cca gga gca      7524
Phe Lys  Ala Met Gly Val Gln  Thr Ile Pro Leu Glu  Pro Gly Ala
    2495                 2500                 2505

cgg act  gtt gcg caa atc att  ttg gca agt tcg cca  ccg caa tcg      7569
Arg Thr  Val Ala Gln Ile Ile  Leu Ala Ser Ser Pro  Pro Gln Ser
    2510                 2515                 2520

ctt ttg  ggg aac tgg ggc ttt  cca gcc acc aaa ccg  cta caa cgc      7614
Leu Leu  Gly Asn Trp Gly Phe  Pro Ala Thr Lys Pro  Leu Gln Arg
    2525                 2530                 2535
```

```
tct aat  gta gtc acg ggc aca  ctc tct ccg gaa gag  ata gaa ttc      7659
Ser Asn  Val Val Thr Gly Thr  Leu Ser Pro Glu Glu  Ile Glu Phe
    2540                 2545                 2550

atc gca  gac cac aaa att caa  ggc cgc aag gtg ctt  ccc atg atg      7704
Ile Ala  Asp His Lys Ile Gln  Gly Arg Lys Val Leu  Pro Met Met
    2555                 2560                 2565

gct gca  atc ggg ttc atg gcc  tct att gcg gaa gga  ctc tac ccg      7749
Ala Ala  Ile Gly Phe Met Ala  Ser Ile Ala Glu Gly  Leu Tyr Pro
    2570                 2575                 2580

ggg tac  aat ctg caa ggc gtg  gaa aat gct cag ctc  ttt caa ggc      7794
Gly Tyr  Asn Leu Gln Gly Val  Glu Asn Ala Gln Leu  Phe Gln Gly
    2585                 2590                 2595

ttg act  atc aac caa gag aca  aaa ttt caa atc act  ctc att gag      7839
Leu Thr  Ile Asn Gln Glu Thr  Lys Phe Gln Ile Thr  Leu Ile Glu
    2600                 2605                 2610

gag cac  aac tct gag gaa aac  ctg gat gtc ctg aca  tcc ctt ggt      7884
Glu His  Asn Ser Glu Glu Asn  Leu Asp Val Leu Thr  Ser Leu Gly
    2615                 2620                 2625

gta atg  ttg gaa agc ggg aag  gtg ctt ccc gct tac  cga tgt gtt      7929
Val Met  Leu Glu Ser Gly Lys  Val Leu Pro Ala Tyr  Arg Cys Val
    2630                 2635                 2640

gta tgc  ttg aat aca acc cag  cag cag ccc aag cta  tct cca aaa      7974
Val Cys  Leu Asn Thr Thr Gln  Gln Gln Pro Lys Leu  Ser Pro Lys
    2645                 2650                 2655

att ctt  aac ttg gaa gtt gac  cct gca tgc gag gtt  aac ccc tat      8019
Ile Leu  Asn Leu Glu Val Asp  Pro Ala Cys Glu Val  Asn Pro Tyr
    2660                 2665                 2670

gat gga  aag tcg ttg ttc cac  ggt ccg ctt ttg caa  ttc gtt caa      8064
Asp Gly  Lys Ser Leu Phe His  Gly Pro Leu Leu Gln  Phe Val Gln
    2675                 2680                 2685

caa gtg  ttg cac tca agt acc  aaa ggc ctc gtt gcc  aag tgc cgc      8109
Gln Val  Leu His Ser Ser Thr  Lys Gly Leu Val Ala  Lys Cys Arg
    2690                 2695                 2700

gcg ctt  cca atc aaa gaa gcc  atc cga ggg cca ttt  atc aag caa      8154
Ala Leu  Pro Ile Lys Glu Ala  Ile Arg Gly Pro Phe  Ile Lys Gln
    2705                 2710                 2715

aca ctc  cat gat cca att cta  gac gac gtc att ttt  cag cta atg      8199
Thr Leu  His Asp Pro Ile Leu  Asp Asp Val Ile Phe  Gln Leu Met
    2720                 2725                 2730

ctc gtg  tgg tgt cgt aat gct  cta gga agt gca tcg  cta ccc aac      8244
Leu Val  Trp Cys Arg Asn Ala  Leu Gly Ser Ala Ser  Leu Pro Asn
    2735                 2740                 2745

aga att  gaa aag atg tca tac  ttt ggg aat gtc tca  gaa ggt agc      8289
Arg Ile  Glu Lys Met Ser Tyr  Phe Gly Asn Val Ser  Glu Gly Ser
    2750                 2755                 2760

act ttc  ttt gcc tca gtt aca  cct gtg gga cca aga  gta cca aag      8334
Thr Phe  Phe Ala Ser Val Thr  Pro Val Gly Pro Arg  Val Pro Lys
    2765                 2770                 2775

gat ccc  gtg atc aaa atg cag  ttt ctt ctc caa gat  gaa tcc ggc      8379
Asp Pro  Val Ile Lys Met Gln  Phe Leu Leu Gln Asp  Glu Ser Gly
    2780                 2785                 2790
```

```
aac aca  ttt tca tcg ggg gag  ggc tcg gtt gtg ctt  agt gac gaa      8424
Asn Thr  Phe Ser Ser Gly Glu  Gly Ser Val Val Leu  Ser Asp Glu
    2795                 2800                 2805

ctc gtc  ttt tga                                                  8436
Leu Val  Phe
    2810
```

<210> 39
<211> 2811
<212> PRT
<213> Thraustochytrium sp.

<400> 39

```
Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1               5                  10                  15

His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
            20                  25                  30

Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
        35                  40                  45

Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
    50                  55                  60

Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65                  70                  75                  80

Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
                85                  90                  95

Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro
            100                 105                 110

Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
        115                 120                 125

Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
    130                 135                 140

Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145                 150                 155                 160

Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
                165                 170                 175

Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
            180                 185                 190

Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala
        195                 200                 205
```

```
Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
    210                 215                 220

Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
225                 230                 235                 240

Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
                245                 250                 255

Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
            260                 265                 270

Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
        275                 280                 285

Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp
    290                 295                 300

Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
305                 310                 315                 320

Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
                325                 330                 335

Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
            340                 345                 350

Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
        355                 360                 365

Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
    370                 375                 380

Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385                 390                 395                 400

Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
                405                 410                 415

Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
            420                 425                 430

Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
        435                 440                 445

Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
    450                 455                 460

Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465                 470                 475                 480
```

```
His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
                485                 490                 495

Cys Asn Asp Gln Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys
            500                 505                 510

Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu
        515                 520                 525

Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe
    530                 535                 540

Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile
545                 550                 555                 560

Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro
                565                 570                 575

His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly
            580                 585                 590

Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met
        595                 600                 605

Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val
    610                 615                 620

Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro
625                 630                 635                 640

Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu
                645                 650                 655

Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr
            660                 665                 670

Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe
        675                 680                 685

Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu
    690                 695                 700

Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys
705                 710                 715                 720

Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala
                725                 730                 735

Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val
            740                 745                 750
```

```
Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr
        755                 760                 765

Val Gln Gly Val Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe
    770                 775                 780

Arg Val Leu Pro Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met
785                 790                 795                 800

Gly Pro Ser Glu Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile
                805                 810                 815

Ser Pro Pro Lys Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu
            820                 825                 830

Ser Pro Asn Pro Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val
        835                 840                 845

Gln Phe Glu Glu Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val
    850                 855                 860

Phe Leu Glu Phe Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu
865                 870                 875                 880

Met Phe Pro Ser Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly
                885                 890                 895

Asp Ser Asp Val Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser
            900                 905                 910

Gly Ala Ala Leu Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln
        915                 920                 925

Asp Leu Leu Ile Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala
    930                 935                 940

Ala Thr Tyr Val Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met
945                 950                 955                 960

Glu Asp Ile Lys Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser
                965                 970                 975

Ile Gly Lys Ile Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu
            980                 985                 990

Thr Glu Ile Arg Arg Leu Asn Lys  Glu Leu Glu His Leu  Lys Arg Glu
        995                 1000                1005

Leu Ala  Ala Ala Lys Ala Ser  Val Lys Ser Ala Ser  Lys Ser Ser
    1010                1015                1020
```

```
Lys Glu  Arg Ser Val Leu Ser  Lys His Arg Ala Leu  Leu Gln Asn
    1025                1030                1035

Ile Leu  Gln Asp Tyr Asp Asp  Leu Arg Val Val Pro  Phe Ala Val
    1040                1045                1050

Arg Ser  Val Ala Val Asp Asn  Thr Ala Pro Tyr Ala  Asp Gln Val
    1055                1060                1065

Ser Thr  Pro Ala Ser Glu Arg  Ser Ala Ser Pro Leu  Phe Glu Lys
    1070                1075                1080

Arg Ser  Ser Val Ser Ser Ala  Arg Leu Ala Glu Ala  Glu Ala Ala
    1085                1090                1095

Val Leu  Ser Val Leu Ala Asp  Lys Thr Gly Tyr Asp  Ser Ser Met
    1100                1105                1110

Ile Glu  Met Asp Met Asp Leu  Glu Ser Glu Leu Gly  Val Asp Ser
    1115                1120                1125

Ile Lys  Arg Val Glu Ile Met  Ser Glu Val Gln Thr  Leu Leu Ser
    1130                1135                1140

Val Glu  Val Ser Asp Val Asp  Ala Leu Ser Arg Thr  Lys Thr Val
    1145                1150                1155

Gly Asp  Val Ile Glu Ala Met  Lys Leu Glu Leu Gly  Gly Pro Gln
    1160                1165                1170

Gly Gln  Thr Leu Thr Ala Glu  Ser Ile Arg Gln Pro  Pro Val Ser
    1175                1180                1185

Glu Pro  Ala Val Pro Thr Ser  Ser Ser Ser Ser Ile  Ala Asn Val
    1190                1195                1200

Ser Ser  Ala Arg Leu Ala Glu  Ala Glu Ala Ala Val  Leu Ser Val
    1205                1210                1215

Leu Ala  Asp Lys Thr Gly Tyr  Asp Ser Ser Met Ile  Glu Met Asp
    1220                1225                1230

Met Asp  Leu Glu Ser Glu Leu  Gly Val Asp Ser Ile  Lys Arg Val
    1235                1240                1245

Glu Ile  Met Ser Glu Val Gln  Thr Leu Leu Ser Val  Glu Val Ser
    1250                1255                1260

Asp Val  Asp Ala Leu Ser Arg  Thr Lys Thr Val Gly  Asp Val Ile
    1265                1270                1275
```

```
Glu Ala  Met Lys Leu Glu Leu  Gly Gly Pro Gln Gly  Gln Thr Leu
    1280                 1285                 1290

Thr Ala  Glu Ser Ile Arg Gln  Pro Pro Val Ser Glu  Pro Ala Val
    1295                 1300                 1305

Pro Thr  Ser Ser Ser Ser Ser  Ile Ala Asn Val Ser  Ser Ala Arg
    1310                 1315                 1320

Leu Ala  Glu Ala Glu Ala Ala  Val Leu Ser Val Leu  Ala Asp Lys
    1325                 1330                 1335

Thr Gly  Tyr Asp Ser Ser Met  Ile Glu Met Asp Met  Asp Leu Glu
    1340                 1345                 1350

Ser Glu  Leu Gly Val Asp Ser  Ile Lys Arg Val Glu  Ile Met Ser
    1355                 1360                 1365

Glu Val  Gln Thr Leu Leu Ser  Val Glu Val Ser Asp  Val Asp Ala
    1370                 1375                 1380

Leu Ser  Arg Thr Lys Thr Val  Gly Asp Val Ile Glu  Ala Met Lys
    1385                 1390                 1395

Leu Glu  Leu Gly Gly Pro Gln  Gly Gln Thr Leu Thr  Ala Glu Ser
    1400                 1405                 1410

Ile Arg  Gln Pro Pro Val Ser  Glu Pro Ala Val Pro  Thr Ser Ser
    1415                 1420                 1425

Ser Ser  Ser Ile Ala Asn Val  Leu Ser Ala Arg Leu  Ala Glu Ala
    1430                 1435                 1440

Glu Ala  Ala Val Leu Ser Val  Leu Ala Asp Lys Thr  Gly Tyr Asp
    1445                 1450                 1455

Ser Ser  Met Ile Glu Met Asp  Met Asp Leu Glu Ser  Glu Leu Gly
    1460                 1465                 1470

Val Asp  Ser Ile Lys Arg Val  Glu Ile Met Ser Glu  Val Gln Thr
    1475                 1480                 1485

Leu Leu  Ser Val Glu Val Ser  Asp Val Asp Ala Leu  Ser Arg Thr
    1490                 1495                 1500

Lys Thr  Val Gly Asp Val Ile  Glu Ala Met Lys Leu  Glu Leu Gly
    1505                 1510                 1515

Gly Pro  Gln Gly Gln Thr Leu  Thr Ala Glu Ser Ile  Arg Gln Pro
    1520                 1525                 1530
```

```
Pro Val  Ser Glu Pro Ala Val  Pro Thr Ser Ser Ser  Ser Ser Ile
    1535                1540                1545

Ala Asn  Val Ser Ser Ala Arg  Leu Ala Glu Ala Glu  Ala Ala Val
    1550                1555                1560

Leu Ser  Val Leu Ala Asp Lys  Thr Gly Tyr Asp Ser  Ser Met Ile
    1565                1570                1575

Glu Met  Asp Met Asp Leu Glu  Ser Glu Leu Gly Val  Asp Ser Ile
    1580                1585                1590

Lys Arg  Val Glu Ile Met Ser  Glu Val Gln Thr Leu  Leu Ser Val
    1595                1600                1605

Glu Val  Ser Asp Val Asp Ala  Leu Ser Arg Thr Lys  Thr Val Gly
    1610                1615                1620

Asp Val  Ile Glu Ala Met Lys  Leu Glu Leu Gly Gly  Pro Gln Gly
    1625                1630                1635

Gln Thr  Leu Thr Ser Glu Pro  Ile His Gln Pro Pro  Val Ser Glu
    1640                1645                1650

Pro Ala  Val Pro Thr Ser Ser  Ser Ser Ser Ile Ala  Asn Val Ser
    1655                1660                1665

Ser Ala  Arg Leu Ala Glu Ala  Glu Ala Ala Val Leu  Ser Val Leu
    1670                1675                1680

Ala Asp  Lys Thr Gly Tyr Asp  Ser Ser Met Ile Glu  Met Asp Met
    1685                1690                1695

Asp Leu  Glu Ser Glu Leu Gly  Val Asp Ser Ile Lys  Arg Val Glu
    1700                1705                1710

Ile Met  Ser Glu Val Gln Thr  Leu Leu Ser Val Glu  Val Ser Asp
    1715                1720                1725

Val Asp  Ala Leu Ser Arg Thr  Lys Thr Val Gly Asp  Val Ile Glu
    1730                1735                1740

Ala Met  Lys Met Glu Leu Gly  Gly Pro Gln Gly Gln  Thr Leu Thr
    1745                1750                1755

Ala Glu  Ser Ile Arg Gln Pro  Pro Val Ser Glu Pro  Ala Val Pro
    1760                1765                1770

Thr Ser  Ser Ser Ser Ser Ile  Ala Asn Val Ser Ser  Ala Arg Leu
    1775                1780                1785
```

```
Ala Glu  Ala Glu Ala Ala Val  Leu Ser Val Leu Ala  Asp Lys Thr
    1790                 1795                 1800

Gly Tyr  Asp Ser Ser Met Ile  Glu Met Asp Met Asp  Leu Glu Ser
    1805                 1810                 1815

Glu Leu  Gly Val Asp Ser Ile  Lys Arg Val Glu Ile  Met Ser Glu
    1820                 1825                 1830

Val Gln  Ala Leu Leu Ser Val  Glu Val Ser Asp Val  Asp Ala Leu
    1835                 1840                 1845

Ser Arg  Thr Lys Thr Val Gly  Asp Val Ile Glu Ala  Met Lys Met
    1850                 1855                 1860

Glu Leu  Gly Gly Pro Gln Gly  Gln Thr Leu Thr Ala  Glu Ser Ile
    1865                 1870                 1875

Arg Glu  Pro Pro Val Ser Glu  Pro Ala Val Pro Thr  Ser Ser Ser
    1880                 1885                 1890

Ser Ser  Ile Ala Asn Val Ser  Ser Ala Arg Leu Ala  Glu Ala Glu
    1895                 1900                 1905

Ala Ala  Val Leu Ser Val Leu  Ala Asp Lys Thr Gly  Tyr Asp Ser
    1910                 1915                 1920

Ser Met  Ile Glu Met Asp Met  Asp Leu Glu Ser Glu  Leu Gly Val
    1925                 1930                 1935

Asp Ser  Ile Lys Arg Val Glu  Ile Met Ser Glu Val  Gln Thr Leu
    1940                 1945                 1950

Leu Ser  Val Glu Val Ser Asp  Val Asp Ala Leu Ser  Arg Thr Lys
    1955                 1960                 1965

Thr Val  Gly Asp Val Ile Glu  Ala Met Lys Leu Glu  Leu Gly Glu
    1970                 1975                 1980

Ser Ser  Ser Ile Glu Thr Leu  Asn Cys Thr Glu Val  Glu His Thr
    1985                 1990                 1995

Ser Tyr  Lys Ser Val Lys Ala  Ser Gly Cys Glu Asn  Val Asp Thr
    2000                 2005                 2010

Arg Phe  Ala Lys Val Val Gln  Ile Ser Leu Pro Ser  Lys Leu Lys
    2015                 2020                 2025

Ser Thr  Val Ser His Asp Arg  Pro Val Ile Val Val  Asp Asp Gly
    2030                 2035                 2040
```

```
Thr Pro  Leu Thr Thr Glu Leu  Cys Lys Ile Leu Gly  Gly Asn Ile
    2045                2050                2055

Val Val  Leu Ser Tyr Gln Gly  Lys Pro Ala Gly Pro  Arg Gly Val
    2060                2065                2070

Glu Val  Pro Asp Leu Ser Glu  Glu Ala Leu Ile Gln  Ala Leu Ala
    2075                2080                2085

Leu Ile  Arg Ser Thr Tyr Gly  Val Pro Ile Gly Phe  Ile Cys Gln
    2090                2095                2100

Gln Val  Ser Asn Val Ser Thr  Lys Ala Gln Leu Cys  Trp Ala Leu
    2105                2110                2115

Leu Ala  Ala Lys His Leu Lys  Lys Asp Leu Asn Ala  Val Leu Pro
    2120                2125                2130

Asp Ser  Arg Ser Phe Phe Val  Gly Val Val Arg Leu  Asn Gly Lys
    2135                2140                2145

Leu Gly  Thr Phe Glu Asn Ile  Ser Asp Phe Ser Lys  Phe Asp Leu
    2150                2155                2160

Thr Lys  Ala Leu Asp Tyr Gly  Gln Arg Gly Ser Leu  Leu Gly Leu
    2165                2170                2175

Cys Lys  Ser Leu Asp Leu Glu  Trp Glu Gln Val Phe  Cys Arg Gly
    2180                2185                2190

Ile Asp  Leu Ala Cys Asp Leu  Met Pro Leu Gln Ala  Ala Arg Ile
    2195                2200                2205

Leu Arg  Asn Glu Leu Gln Cys  Pro Asn Met Arg Leu  Arg Glu Val
    2210                2215                2220

Gly Tyr  Asp Ile Ser Gly Ala  Arg Tyr Thr Ile Ser  Thr Asp Asp
    2225                2230                2235

Leu Leu  Cys Gly Pro Ser Lys  Ala Lys Val Glu Ala  Ala Asp Leu
    2240                2245                2250

Phe Leu  Val Thr Gly Gly Ala  Arg Gly Ile Thr Pro  His Cys Val
    2255                2260                2265

Arg Glu  Ile Ala Ser Arg Ser  Pro Gly Thr Thr Phe  Val Leu Val
    2270                2275                2280

Gly Arg  Ser Glu Met Ser Asp  Glu Pro Asp Trp Ala  Val Gly His
    2285                2290                2295
```

```
Tyr Asn  Lys Asp Leu Asp Gln  Ser Thr Met Lys His  Leu Lys Ala
    2300                 2305                 2310

Thr His  Ala Ala Gly Gly Val  Lys Pro Thr Pro Lys  Ala His Arg
    2315                 2320                 2325

Ala Leu  Val Asn Arg Val Thr  Gly Ser Arg Glu Val  Arg Glu Ser
    2330                 2335                 2340

Leu Arg  Ala Ile Gln Glu Ala  Gly Ala Asn Val Glu  Tyr Ile Ala
    2345                 2350                 2355

Cys Asp  Val Ser Asp Glu Asn  Lys Val Arg Gln Leu  Val Gln Arg
    2360                 2365                 2370

Val Glu  Gln Lys Tyr Gly Cys  Glu Ile Thr Gly Ile  Trp His Ala
    2375                 2380                 2385

Ser Gly  Val Leu Arg Asp Lys  Leu Val Glu Gln Lys  Thr Thr Asp
    2390                 2395                 2400

Asp Phe  Glu Ala Val Phe Gly  Thr Lys Val Thr Gly  Leu Val Asn
    2405                 2410                 2415

Ile Val  Ser Gln Val Asn Met  Ser Lys Leu Arg His  Phe Ile Leu
    2420                 2425                 2430

Phe Ser  Ser Leu Ala Gly Phe  His Gly Asn Lys Gly  Gln Thr Asp
    2435                 2440                 2445

Tyr Ala  Ile Ala Asn Glu Ala  Leu Asn Lys Ile Ala  His Thr Leu
    2450                 2455                 2460

Ser Ala  Phe Leu Pro Lys Leu  Asn Ala Lys Val Leu  Asp Phe Gly
    2465                 2470                 2475

Pro Trp  Val Gly Ser Gly Met  Val Thr Glu Thr Leu  Glu Lys His
    2480                 2485                 2490

Phe Lys  Ala Met Gly Val Gln  Thr Ile Pro Leu Glu  Pro Gly Ala
    2495                 2500                 2505

Arg Thr  Val Ala Gln Ile Ile  Leu Ala Ser Ser Pro  Pro Gln Ser
    2510                 2515                 2520

Leu Leu  Gly Asn Trp Gly Phe  Pro Ala Thr Lys Pro  Leu Gln Arg
    2525                 2530                 2535

Ser Asn  Val Val Thr Gly Thr  Leu Ser Pro Glu Glu  Ile Glu Phe
    2540                 2545                 2550
```

```
Ile Ala  Asp His Lys Ile Gln  Gly Arg Lys Val Leu  Pro Met Met
    2555                2560                2565

Ala Ala  Ile Gly Phe Met Ala  Ser Ile Ala Glu Gly  Leu Tyr Pro
    2570                2575                2580

Gly Tyr  Asn Leu Gln Gly Val  Glu Asn Ala Gln Leu  Phe Gln Gly
    2585                2590                2595

Leu Thr  Ile Asn Gln Glu Thr  Lys Phe Gln Ile Thr  Leu Ile Glu
    2600                2605                2610

Glu His  Asn Ser Glu Glu Asn  Leu Asp Val Leu Thr  Ser Leu Gly
    2615                2620                2625

Val Met  Leu Glu Ser Gly Lys  Val Leu Pro Ala Tyr  Arg Cys Val
    2630                2635                2640

Val Cys  Leu Asn Thr Thr Gln  Gln Gln Pro Lys Leu  Ser Pro Lys
    2645                2650                2655

Ile Leu  Asn Leu Glu Val Asp  Pro Ala Cys Glu Val  Asn Pro Tyr
    2660                2665                2670

Asp Gly  Lys Ser Leu Phe His  Gly Pro Leu Leu Gln  Phe Val Gln
    2675                2680                2685

Gln Val  Leu His Ser Ser Thr  Lys Gly Leu Val Ala  Lys Cys Arg
    2690                2695                2700

Ala Leu  Pro Ile Lys Glu Ala  Ile Arg Gly Pro Phe  Ile Lys Gln
    2705                2710                2715

Thr Leu  His Asp Pro Ile Leu  Asp Asp Val Ile Phe  Gln Leu Met
    2720                2725                2730

Leu Val  Trp Cys Arg Asn Ala  Leu Gly Ser Ala Ser  Leu Pro Asn
    2735                2740                2745

Arg Ile  Glu Lys Met Ser Tyr  Phe Gly Asn Val Ser  Glu Gly Ser
    2750                2755                2760

Thr Phe  Phe Ala Ser Val Thr  Pro Val Gly Pro Arg  Val Pro Lys
    2765                2770                2775

Asp Pro  Val Ile Lys Met Gln  Phe Leu Leu Gln Asp  Glu Ser Gly
    2780                2785                2790

Asn Thr  Phe Ser Ser Gly Glu  Gly Ser Val Val Leu  Ser Asp Glu
    2795                2800                2805
```

Leu Val Phe
2810

<210> 40
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 40

```
atg aag gac atg gaa gat aga cgg gtc gct att gtg ggc atg tca gct       48
Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1               5                   10                  15

cac ttg cct tgt ggg aca gat gtg aag gaa tca tgg cag gct att cgc       96
His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
            20                  25                  30

gat gga atc gac tgt cta agt gac cta ccc gcg gat cgt ctc gac gtt      144
Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
        35                  40                  45

aca gct tac tac aat ccc aac aaa gcc acg aaa gac aag atc tac tgc      192
Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
    50                  55                  60

aaa cgg ggt ggc ttc atc ccg aac tat gac ttc gac ccc cgc gaa ttt      240
Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65                  70                  75                  80

ggg ctc aac atg ttt caa atg gaa gac tct gat gcg aat cag aca ctt      288
Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
                85                  90                  95

acc ttg ctc aaa gtc aaa caa gct ctc gaa gat gca agc ata gag cct      336
Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro
            100                 105                 110

ttc acc aag gag aag aag aac att gga tgt gtt tta ggt att ggt ggg      384
Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
        115                 120                 125

ggc caa aag gcg agt cat gag ttc tac tct cgt ctc aac tac gtt gtc      432
Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
    130                 135                 140

gtt gaa aag gta ctt cgg aaa atg ggt tta cca gat gct gat gtt gaa      480
Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145                 150                 155                 160

gaa gct gtg gag aaa tac aag gca aat ttt ccc gag tgg cgc cta gac      528
Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
                165                 170                 175

tct ttc cct ggg ttt ctt ggg aat gta acg gct ggt cgg tgc agt aac      576
Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
            180                 185                 190

acc ttc aac atg gaa ggt atg aac tgc gtt gtg gat gct gca tgt gcc      624
Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala
        195                 200                 205

agt tct cta att gca atc aag gtt gca gtt gaa gag cta ctc ttt ggt      672
```

```
Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
    210                 215                 220

gac tgt gac acc atg att gca ggt gcc acc tgc acg gac aat tca ctt      720
Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
225                 230                 235                 240

ggc atg tac atg gcc ttc tct aaa acg cca gtt ttt tct act gac cca      768
Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
                245                 250                 255

agt gtc cgc gcg tat gat gag aaa aca aaa ggg atg cta att gga gaa      816
Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
            260                 265                 270

ggt tca gca atg ttc gtt ctt aaa cgc tat gcg gat gcc gta cgt gat      864
Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
        275                 280                 285

ggc gac aca att cac gcg gtt ctg cgt tct tgc tct tcg tct agt gat      912
Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp
    290                 295                 300

gga aaa gcg gca gga att tat act cct act ata tct gga caa gaa gaa      960
Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
305                 310                 315                 320

gct ttg cgt cga gcg tat gcc cgt gcg ggg gta tgt cca tct acg atc     1008
Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
                325                 330                 335

ggg ctt gtt gag ggt cac ggg aca ggg acc cct gtt gga gat cgc att     1056
Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
            340                 345                 350

gag tta aca gct ctg cgg aac ttg ttt gac aaa gct ttt ggt agc aag     1104
Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
        355                 360                 365

aag gaa caa ata gca gtt ggc agc ata aag tct cag ata ggt cac ctg     1152
Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
    370                 375                 380

aaa tct gtt gcc ggc ttt gcc ggc ttg gtc aaa gct gtg ctt gcg ctt     1200
Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385                 390                 395                 400

aaa cac aaa acg ctc cca ggt tcg att aat gtc gac cag cca cct ttg     1248
Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
                405                 410                 415

ttg tat gac ggt act caa att caa gac tct tct tta tat atc aac aag     1296
Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
            420                 425                 430

aca aat aga cca tgg ttt acg caa aac aag ctt ccg cgt cgg gct ggt     1344
Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
        435                 440                 445

gtc tca agt ttt gga ttt gga ggt gca aac tac cac gcg gtt ctg gaa     1392
Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
    450                 455                 460

gaa ttc gag ccc gag cat gaa aaa cca tac cgc ctc aat act gtt gga     1440
Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465                 470                 475                 480

cat cct gtc ctc ttg tac gct ccg tct gtg gaa gcc ctc aaa gta ctt     1488
```

```
His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
                485                 490                 495

tgc aac gac cag                                                 1500
Cys Asn Asp Gln
            500


<210>  41
<211>  500
<212>  PRT
<213>  Thraustochytrium sp.

<400>  41

Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1               5                   10                  15


His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
            20                  25                  30


Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
        35                  40                  45


Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
    50                  55                  60


Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65                  70                  75                  80


Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
                85                  90                  95


Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro
            100                 105                 110


Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
        115                 120                 125


Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
    130                 135                 140


Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145                 150                 155                 160


Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
                165                 170                 175


Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
            180                 185                 190


Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala
        195                 200                 205


Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
```

210 215 220

Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
225 230 235 240

Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
245 250 255

Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
260 265 270

Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
275 280 285

Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp
290 295 300

Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
305 310 315 320

Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
325 330 335

Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
340 345 350

Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
355 360 365

Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
370 375 380

Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385 390 395 400

Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
405 410 415

Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
420 425 430

Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
435 440 445

Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
450 455 460

Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465 470 475 480

His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu

```
485                 490                 495

Cys Asn Asp Gln
            500


<210>  42
<211>  1500
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<400>  42
ctt gcg gag ctc aca att gca ttg gaa gag gca aaa aca cat aaa aat       48
Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys Thr His Lys Asn
1               5                   10                  15

gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa ttt cag ctc caa       96
Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu Phe Gln Leu Gln
            20                  25                  30

gga agc tgt cct cca gaa aat ccg aga gta gga ttt tta gca aca ctg      144
Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe Leu Ala Thr Leu
        35                  40                  45

cct act tca aat atc att gtc gcg ctt aag gca att ctc gcg cag ctt      192
Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile Leu Ala Gln Leu
    50                  55                  60

gat gca aaa cca gat gcg aag aaa tgg gat ttg cct cat aaa aag gct      240
Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro His Lys Lys Ala
65                  70                  75                  80

ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc tct gtt gct gcg      288
Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala
                85                  90                  95

ctc ttc gca gga cag ggt acc cag tac tta aac atg ttc tct gat gtg      336
Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val
            100                 105                 110

gca atg aac tgg cca ccg ttc cgt gac agc att gtc gca atg gaa gaa      384
Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val Ala Met Glu Glu
        115                 120                 125

gct caa act gag gta ttt gag ggc caa gtt gaa cca att agc aaa gtt      432
Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro Ile Ser Lys Val
    130                 135                 140

ctg ttt cca cga gag cgc tat gca tcc gaa agt gaa cag ggg aat gaa      480
Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu Gln Gly Asn Glu
145                 150                 155                 160

ctt ctt tgc tta aca gag tac tct cag cca act acg ata gca gcc gca      528
Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr Ile Ala Ala Ala
                165                 170                 175

gta ggg gcc ttc gat att ttc aaa gcg gct ggc ttt aag cca gac atg      576
Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe Lys Pro Asp Met
            180                 185                 190

gtt gga ggg cat tca ctt ggc gaa ttt gct gct ttg tac gcg gct ggg      624
Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly
```

```
        195                 200                 205

tcc att tcg cgt gac gac ctg tac aag ctt gtg tgc aaa cgg gca aag      672
Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys Lys Arg Ala Lys
    210                 215                 220

gca atg gcg aac gct agt gac gga gct atg gca gca gtg att ggc cca      720
Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala Val Ile Gly Pro
225                 230                 235                 240

gat gca cgt cta gtt acg cca caa aat agt gac gtt tat gtc gca aac      768
Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val Tyr Val Ala Asn
                245                 250                 255

ttc aac tcc gca act caa gta gtc atc agt ggc act gtt caa ggt gtg      816
Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr Val Gln Gly Val
            260                 265                 270

aaa gaa gag tcg aaa ttg ctc att tca aag ggg ttc cgc gta ctg cca      864
Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe Arg Val Leu Pro
        275                 280                 285

ctt aaa tgc cag ggc gcc ttc cat tct cct ttg atg ggg cct tct gag      912
Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met Gly Pro Ser Glu
    290                 295                 300

gat agt ttc aaa tca ctt gtg gag act tgt acc atc tcg ccg cca aaa      960
Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile Ser Pro Pro Lys
305                 310                 315                 320

aat gtg aaa ttc ttt tgc aat gtt agt ggc aag gaa agc cca aac cca     1008
Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu Ser Pro Asn Pro
                325                 330                 335

aaa cag acc ctc aag tca cac atg acg tct agc gtt cag ttc gag gag     1056
Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val Gln Phe Glu Glu
            340                 345                 350

cag att cgt aac atg tac gat gcc gga gca cgt gtt ttt ctg gag ttt     1104
Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val Phe Leu Glu Phe
        355                 360                 365

gga ccc cgc caa gtc ctt gca aag ctt atc gcg gaa atg ttt ccc tcg     1152
Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu Met Phe Pro Ser
    370                 375                 380

tgt aca gct atc agc gtt aac ccc gcg agc agt ggt gac agt gac gtg     1200
Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly Asp Ser Asp Val
385                 390                 395                 400

caa ctc cgc ctc gcc gcc gta aaa ttc gcg gtc tcg ggt gca gcc ctt     1248
Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser Gly Ala Ala Leu
                405                 410                 415

agc acc ttt gat cca tgg gag tat cgc aag cca caa gat ctt ctt att     1296
Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln Asp Leu Leu Ile
            420                 425                 430

cga aaa cca cga aaa act gcc ctt gtt cta tca gca gca aca tat gtt     1344
Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala Ala Thr Tyr Val
        435                 440                 445

tcc cca aag act ctt gca gaa cgt aaa aag gct atg gaa gat atc aag     1392
Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met Glu Asp Ile Lys
    450                 455                 460

cta gta tcc att aca cca aga gat agt atg gta tca att gga aaa atc     1440
Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser Ile Gly Lys Ile
```

```
465                 470                 475                 480

gcg caa gaa gta cgg aca gct aaa cag cct tta gaa acc gaa att cga     1488
Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu Thr Glu Ile Arg
                485                 490                 495

aga ctc aac aaa                                                     1500
Arg Leu Asn Lys
            500


<210>  43
<211>  500
<212>  PRT
<213>  Thraustochytrium sp.

<400>  43

Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys Thr His Lys Asn
1               5                   10                  15


Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu Phe Gln Leu Gln
            20                  25                  30


Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe Leu Ala Thr Leu
        35                  40                  45


Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile Leu Ala Gln Leu
    50                  55                  60


Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro His Lys Lys Ala
65                  70                  75                  80


Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala
                85                  90                  95


Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val
            100                 105                 110


Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val Ala Met Glu Glu
        115                 120                 125


Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro Ile Ser Lys Val
    130                 135                 140


Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu Gln Gly Asn Glu
145                 150                 155                 160


Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr Ile Ala Ala Ala
                165                 170                 175


Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe Lys Pro Asp Met
            180                 185                 190


Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly
        195                 200                 205
```

```
Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys Lys Arg Ala Lys
    210                 215                 220

Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala Val Ile Gly Pro
225                 230                 235                 240

Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val Tyr Val Ala Asn
                245                 250                 255

Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr Val Gln Gly Val
            260                 265                 270

Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe Arg Val Leu Pro
        275                 280                 285

Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met Gly Pro Ser Glu
    290                 295                 300

Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile Ser Pro Pro Lys
305                 310                 315                 320

Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu Ser Pro Asn Pro
                325                 330                 335

Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val Gln Phe Glu Glu
            340                 345                 350

Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val Phe Leu Glu Phe
        355                 360                 365

Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu Met Phe Pro Ser
    370                 375                 380

Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly Asp Ser Asp Val
385                 390                 395                 400

Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser Gly Ala Ala Leu
                405                 410                 415

Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln Asp Leu Leu Ile
            420                 425                 430

Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala Ala Thr Tyr Val
        435                 440                 445

Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met Glu Asp Ile Lys
    450                 455                 460

Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser Ile Gly Lys Ile
465                 470                 475                 480
```

```
Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu Thr Glu Ile Arg
                485                 490                 495


Arg Leu Asn Lys
            500


<210>  44
<211>  351
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(351)

<400>  44
tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc       48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1               5                   10                  15

agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg       96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
            20                  25                  30

agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg      144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
        35                  40                  45

gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg      192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
    50                  55                  60

gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac      240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65                  70                  75                  80

gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg      288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
                85                  90                  95

atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa      336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
            100                 105                 110

tcg atc cgt cag cca                                                 351
Ser Ile Arg Gln Pro
        115


<210>  45
<211>  117
<212>  PRT
<213>  Thraustochytrium sp.

<400>  45

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1               5                   10                  15


Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
            20                  25                  30
```

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
        35   40   45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
    50   55   60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65   70   75   80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
                85   90   95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
            100   105   110

Ser Ile Arg Gln Pro
        115

<210> 46
<211> 5
<212> PRT
<213> Thraustochytrium sp.

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Xaa = any amino acid

<400> 46

Leu Gly Xaa Asp Ser
1   5

<210> 47
<211> 2790
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(2790)

<400> 47
tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc   48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1   5   10   15

agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg   96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
            20   25   30

agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg   144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
        35   40   45

gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg   192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
    50   55   60

```
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac      240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65                  70                  75                  80

gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg      288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
                85                  90                  95

atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa      336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
            100                 105                 110

tcg atc cgt cag cca ccg gtg tcc gag cct gct gta ccg acc tca tcg      384
Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
        115                 120                 125

tca agc agt att gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa      432
Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
    130                 135                 140

gct gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca      480
Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
145                 150                 155                 160

atg atc gag atg gac atg gac ctg gag agc gag ctt ggc gtt gat agc      528
Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
                165                 170                 175

atc aaa cgc gtg gag atc atg agc gag gtt caa acg ctg ctc agc gtg      576
Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
            180                 185                 190

gaa gtc tcc gac gtt gac gct ctg tca aga act aag act gtt ggc gac      624
Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
        195                 200                 205

gtc atc gag gcg atg aag ctg gaa ctc ggt gga ccc caa ggc cag act      672
Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr
    210                 215                 220

ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta      720
Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
225                 230                 235                 240

ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc      768
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
                245                 250                 255

gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca ggc      816
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
            260                 265                 270

tac gac agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt      864
Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
        275                 280                 285

ggc gtc gac agc atc aaa cgc gtg gag atc atg agc gag gtt caa acg      912
Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr
    290                 295                 300

ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag      960
Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
305                 310                 315                 320

act gtt ggc gac gtc atc gag gcg atg aag ctg gaa ctc ggt gga ccc     1008
Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro
                325                 330                 335
```

```
caa ggc cag act ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tcc     1056
Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser
            340                 345                 350

gag cct gct gta ccg acc tca tcg tca agc agt att gct aat gtt ttg     1104
Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile Ala Asn Val Leu
        355                 360                 365

tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg agc gtt ctc gca     1152
Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
    370                 375                 380

gac aag aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg     1200
Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
385                 390                 395                 400

gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg gag atc atg agc     1248
Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
                405                 410                 415

gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg     1296
Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
            420                 425                 430

tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa     1344
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
        435                 440                 445

ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa tcg atc cgt cag     1392
Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln
    450                 455                 460

cca ccg gtg tct gag cct gct gta ccg acc tca tcg tca agc agt att     1440
Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile
465                 470                 475                 480

gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg     1488
Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
                485                 490                 495

agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg     1536
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
            500                 505                 510

gac atg gac ctg gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg     1584
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
        515                 520                 525

gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac     1632
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
    530                 535                 540

gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg     1680
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
545                 550                 555                 560

atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc tct gaa     1728
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu
                565                 570                 575

ccg atc cat cag cca cca gtg tcc gag cct gct gta ccg acc tca tcg     1776
Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
            580                 585                 590

tca agc agt att gct aat gtt tct tca gca cgc ctc gct gaa gct gaa     1824
Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
        595                 600                 605
```

```
gcc gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca     1872
Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
    610                 615                 620

atg atc gag atg gac atg gac ctg gag agc gag ctt ggc gtt gat agc     1920
Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
625                 630                 635                 640

atc aaa cgc gtg gaa atc atg agc gag gtt caa acg ctg ctc agc gtg     1968
Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
                645                 650                 655

gaa gtc tcc gac gtt gac gct ctg tca aga acc aag act gtt ggc gac     2016
Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
            660                 665                 670

gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc caa ggc cag act     2064
Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr
        675                 680                 685

ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta     2112
Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
    690                 695                 700

ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc     2160
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
705                 710                 715                 720

gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca ggc     2208
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
                725                 730                 735

tac gac agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt     2256
Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
            740                 745                 750

ggc gtt gat agc atc aaa cgc gtg gag atc atg agc gag gtt caa gcg     2304
Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala
        755                 760                 765

ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag     2352
Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
    770                 775                 780

act gtt ggc gac gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc     2400
Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro
785                 790                 795                 800

caa ggc cag act ttg acc gca gaa tcg atc cgt gag cca ccg gtg tct     2448
Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser
                805                 810                 815

gag cct gct gta ccg acc tca tcg tca agt agt atc gct aat gtt tct     2496
Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser
            820                 825                 830

tca gct cgc ctc gct gaa gct gaa gcc gcg gta ctg agc gtt ctc gca     2544
Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
        835                 840                 845

gac aag aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg     2592
Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
    850                 855                 860

gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg gag atc atg agc     2640
Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
865                 870                 875                 880
```

```
gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg     2688
Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
                885                 890                 895

tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa     2736
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
            900                 905                 910

ctt ggg gaa tca tca agt att gag act ctc aat tgt acc gag gtt gag     2784
Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
        915                 920                 925

cac acg                                                            2790
His Thr
    930
```

<210> 48
<211> 930
<212> PRT
<213> Thraustochytrium sp.

<400> 48

```
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1               5                   10                  15

Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
            20                  25                  30

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
        35                  40                  45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
    50                  55                  60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65                  70                  75                  80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
                85                  90                  95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
            100                 105                 110

Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
        115                 120                 125

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
    130                 135                 140

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
145                 150                 155                 160

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
                165                 170                 175
```

```
Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
            180                 185                 190

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
        195                 200                 205

Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr
    210                 215                 220

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
225                 230                 235                 240

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
                245                 250                 255

Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
            260                 265                 270

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
        275                 280                 285

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr
    290                 295                 300

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
305                 310                 315                 320

Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro
                325                 330                 335

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser
            340                 345                 350

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Leu
        355                 360                 365

Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
    370                 375                 380

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
385                 390                 395                 400

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
                405                 410                 415

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
            420                 425                 430

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
        435                 440                 445
```

```
Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln
    450                 455                 460

Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile
465                 470                 475                 480

Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
                485                 490                 495

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
            500                 505                 510

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
        515                 520                 525

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
    530                 535                 540

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
545                 550                 555                 560

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu
                565                 570                 575

Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
            580                 585                 590

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
        595                 600                 605

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
    610                 615                 620

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
625                 630                 635                 640

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
                645                 650                 655

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
            660                 665                 670

Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr
        675                 680                 685

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
    690                 695                 700

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
705                 710                 715                 720
```

```
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
                725                 730                 735

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
            740                 745                 750

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala
        755                 760                 765

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
    770                 775                 780

Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro
785                 790                 795                 800

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser
                805                 810                 815

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser
            820                 825                 830

Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
        835                 840                 845

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
    850                 855                 860

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
865                 870                 875                 880

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
                885                 890                 895

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
            900                 905                 910

Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
        915                 920                 925

His Thr
    930
```

<210> 49
<211> 2433
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(2433)

<400> 49

```
aaa agt gtc aag gct tca ggg tgt gag aat gta gat acc cgt ttc gct       48
Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr Arg Phe Ala
1               5                  10                  15

aag gtt gta caa atc tcg ctt cct agc aag ctg aaa tcc act gtg tcg       96
Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys Ser Thr Val Ser
            20                  25                  30

cac gat cga cct gta att gtt gta gat gat gga acg ccc tta acc acg      144
His Asp Arg Pro Val Ile Val Val Asp Asp Gly Thr Pro Leu Thr Thr
        35                  40                  45

gag ctt tgt aaa att ctt ggg ggt aat att gtg gtt ctc tct tat caa      192
Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile Val Val Leu Ser Tyr Gln
    50                  55                  60

ggg aag ccc gct ggt cca cgg gga gtc gag gtg cca gat ctt tcc gag      240
Gly Lys Pro Ala Gly Pro Arg Gly Val Glu Val Pro Asp Leu Ser Glu
65                  70                  75                  80

gaa gcc cta att caa gct ctt gca ttg att cgg tct aca tat gga gtt      288
Glu Ala Leu Ile Gln Ala Leu Ala Leu Ile Arg Ser Thr Tyr Gly Val
                85                  90                  95

cca att ggt ttt att tgt cag caa gtg tct aat gtg agc acc aag gca      336
Pro Ile Gly Phe Ile Cys Gln Gln Val Ser Asn Val Ser Thr Lys Ala
            100                 105                 110

cag ctt tgt tgg gca ctc ctc gca gcg aag cat ctc aag aag gat ttg      384
Gln Leu Cys Trp Ala Leu Leu Ala Ala Lys His Leu Lys Lys Asp Leu
        115                 120                 125

aat gct gtc tta ccc gat tca aga tcc ttc ttc gtc gga gtt gta cgc      432
Asn Ala Val Leu Pro Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg
    130                 135                 140

ttg aac ggg aaa ctt gga act ttc gaa aac atc agc gac ttc tct aaa      480
Leu Asn Gly Lys Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys
145                 150                 155                 160

ttt gat ttg acg aaa gcc cta gat tac gga cag cgt ggt tct ctc tta      528
Phe Asp Leu Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu
                165                 170                 175

ggc ctg tgc aag tca cta gac tta gaa tgg gaa cag gtg ttt tgc cgt      576
Gly Leu Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg
            180                 185                 190

gga ata gat ctt gcg tgt gat ctt atg cca ctc cag gcc gca agg ata      624
Gly Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile
        195                 200                 205

ctc aga aat gag ctt cag tgt ccc aat atg cgc ctt cgc gag gtt ggg      672
Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val Gly
    210                 215                 220

tac gat att tct ggc gcc agg tac acc att tca acc gat gac ctg cta      720
Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp Leu Leu
225                 230                 235                 240

tgt gga ccc tcg aag gct aaa gta gag gcc gca gac ttg ttt ctt gtg      768
Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu Phe Leu Val
                245                 250                 255

aca ggt ggc gca cga ggt att aca cct cat tgt gtt cgt gag att gca      816
Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val Arg Glu Ile Ala
            260                 265                 270
```

```
agt cga tcc ccc gga acc aca ttt gtg ctg gtt gga aga agc gaa atg      864
Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val Gly Arg Ser Glu Met
        275                 280                 285

tcc gac gag cct gac tgg gct gtt ggc cac tac aat aaa gac ctg gac      912
Ser Asp Glu Pro Asp Trp Ala Val Gly His Tyr Asn Lys Asp Leu Asp
    290                 295                 300

caa agc aca atg aaa cac ttg aaa gca acg cat gct gct gga ggg gta      960
Gln Ser Thr Met Lys His Leu Lys Ala Thr His Ala Ala Gly Gly Val
305                 310                 315                 320

aaa cct acg cct aaa gca cat cgt gca ctt gtg aac agg gtc act ggc     1008
Lys Pro Thr Pro Lys Ala His Arg Ala Leu Val Asn Arg Val Thr Gly
                325                 330                 335

tca cgg gag gta cga gaa tct ctt aga gca atc cag gag gca ggg gca     1056
Ser Arg Glu Val Arg Glu Ser Leu Arg Ala Ile Gln Glu Ala Gly Ala
            340                 345                 350

aat gtc gaa tat atc gcc tgt gat gtt tcg gat gaa aac aag gtc cgc     1104
Asn Val Glu Tyr Ile Ala Cys Asp Val Ser Asp Glu Asn Lys Val Arg
        355                 360                 365

caa ctt gtg caa aga gtg gag caa aag tat ggc tgt gaa ata act ggg     1152
Gln Leu Val Gln Arg Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly
    370                 375                 380

att tgg cat gca agc ggg gtt ctt cgt gac aaa ctt gtc gag caa aag     1200
Ile Trp His Ala Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys
385                 390                 395                 400

act aca gac gac ttt gag gca gtt ttt ggg acc aag gtg act ggc ctt     1248
Thr Thr Asp Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu
                405                 410                 415

gta aac atc gtg tca caa gtc aat atg tct aag cta cga cac ttc atc     1296
Val Asn Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile
            420                 425                 430

ctc ttc agt tct ttg gct gga ttt cat ggg aac aag ggc caa acg gat     1344
Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp
        435                 440                 445

tat gca att gct aat gaa gcc ttg aac aaa atc gcg cat act ctc tca     1392
Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu Ser
    450                 455                 460

gcg ttt ttg ccc aaa ctg aat gca aag gtg cta gac ttc ggt ccg tgg     1440
Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly Pro Trp
465                 470                 475                 480

gta ggt tca gga atg gta acc gaa aca ctt gag aag cat ttt aaa gct     1488
Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His Phe Lys Ala
                485                 490                 495

atg ggg gtt cag act att cct ctc gag cca gga gca cgg act gtt gcg     1536
Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala Arg Thr Val Ala
            500                 505                 510

caa atc att ttg gca agt tcg cca ccg caa tcg ctt ttg ggg aac tgg     1584
Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp
        515                 520                 525

ggc ttt cca gcc acc aaa ccg cta caa cgc tct aat gta gtc acg ggc     1632
Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly
    530                 535                 540
```

```
aca ctc tct ccg gaa gag ata gaa ttc atc gca gac cac aaa att caa     1680
Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln
545                 550                 555                 560

ggc cgc aag gtg ctt ccc atg atg gct gca atc ggg ttc atg gcc tct     1728
Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser
                565                 570                 575

att gcg gaa gga ctc tac ccg ggg tac aat ctg caa ggc gtg gaa aat     1776
Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn
            580                 585                 590

gct cag ctc ttt caa ggc ttg act atc aac caa gag aca aaa ttt caa     1824
Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln
        595                 600                 605

atc act ctc att gag gag cac aac tct gag gaa aac ctg gat gtc ctg     1872
Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu
    610                 615                 620

aca tcc ctt ggt gta atg ttg gaa agc ggg aag gtg ctt ccc gct tac     1920
Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr
625                 630                 635                 640

cga tgt gtt gta tgc ttg aat aca acc cag cag cag ccc aag cta tct     1968
Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser
                645                 650                 655

cca aaa att ctt aac ttg gaa gtt gac cct gca tgc gag gtt aac ccc     2016
Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro
            660                 665                 670

tat gat gga aag tcg ttg ttc cac ggt ccg ctt ttg caa ttc gtt caa     2064
Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln
        675                 680                 685

caa gtg ttg cac tca agt acc aaa ggc ctc gtt gcc aag tgc cgc gcg     2112
Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala
    690                 695                 700

ctt cca atc aaa gaa gcc atc cga ggg cca ttt atc aag caa aca ctc     2160
Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu
705                 710                 715                 720

cat gat cca att cta gac gac gtc att ttt cag cta atg ctc gtg tgg     2208
His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp
                725                 730                 735

tgt cgt aat gct cta gga agt gca tcg cta ccc aac aga att gaa aag     2256
Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys
            740                 745                 750

atg tca tac ttt ggg aat gtc tca gaa ggt agc act ttc ttt gcc tca     2304
Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser
        755                 760                 765

gtt aca cct gtg gga cca aga gta cca aag gat ccc gtg atc aaa atg     2352
Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met
    770                 775                 780

cag ttt ctt ctc caa gat gaa tcc ggc aac aca ttt tca tcg ggg gag     2400
Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785                 790                 795                 800

ggc tcg gtt gtg ctt agt gac gaa ctc gtc ttt                         2433
Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
                805                 810
```

<210> 50
<211> 811
<212> PRT
<213> Thraustochytrium sp.

<400> 50

Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr Arg Phe Ala
1 5 10 15

Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys Ser Thr Val Ser
20 25 30

His Asp Arg Pro Val Ile Val Val Asp Asp Gly Thr Pro Leu Thr Thr
35 40 45

Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile Val Val Leu Ser Tyr Gln
50 55 60

Gly Lys Pro Ala Gly Pro Arg Gly Val Glu Val Pro Asp Leu Ser Glu
65 70 75 80

Glu Ala Leu Ile Gln Ala Leu Ala Leu Ile Arg Ser Thr Tyr Gly Val
85 90 95

Pro Ile Gly Phe Ile Cys Gln Gln Val Ser Asn Val Ser Thr Lys Ala
100 105 110

Gln Leu Cys Trp Ala Leu Leu Ala Ala Lys His Leu Lys Lys Asp Leu
115 120 125

Asn Ala Val Leu Pro Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg
130 135 140

Leu Asn Gly Lys Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys
145 150 155 160

Phe Asp Leu Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu
165 170 175

Gly Leu Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg
180 185 190

Gly Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile
195 200 205

Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val Gly
210 215 220

Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp Leu Leu
225 230 235 240

```
Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu Phe Leu Val
                245                 250                 255

Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val Arg Glu Ile Ala
            260                 265                 270

Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val Gly Arg Ser Glu Met
        275                 280                 285

Ser Asp Glu Pro Asp Trp Ala Val Gly His Tyr Asn Lys Asp Leu Asp
    290                 295                 300

Gln Ser Thr Met Lys His Leu Lys Ala Thr His Ala Ala Gly Gly Val
305                 310                 315                 320

Lys Pro Thr Pro Lys Ala His Arg Ala Leu Val Asn Arg Val Thr Gly
                325                 330                 335

Ser Arg Glu Val Arg Glu Ser Leu Arg Ala Ile Gln Glu Ala Gly Ala
            340                 345                 350

Asn Val Glu Tyr Ile Ala Cys Asp Val Ser Asp Glu Asn Lys Val Arg
        355                 360                 365

Gln Leu Val Gln Arg Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly
    370                 375                 380

Ile Trp His Ala Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys
385                 390                 395                 400

Thr Thr Asp Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu
                405                 410                 415

Val Asn Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile
            420                 425                 430

Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp
        435                 440                 445

Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu Ser
    450                 455                 460

Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly Pro Trp
465                 470                 475                 480

Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His Phe Lys Ala
                485                 490                 495

Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala Arg Thr Val Ala
            500                 505                 510
```

```
Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp
        515                 520                 525

Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly
    530                 535                 540

Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln
545                 550                 555                 560

Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser
                565                 570                 575

Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn
            580                 585                 590

Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln
        595                 600                 605

Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu
    610                 615                 620

Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr
625                 630                 635                 640

Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser
                645                 650                 655

Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro
            660                 665                 670

Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln
        675                 680                 685

Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala
    690                 695                 700

Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu
705                 710                 715                 720

His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp
                725                 730                 735

Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys
            740                 745                 750

Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser
        755                 760                 765

Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met
    770                 775                 780
```

```
Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785                 790                 795                 800

Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
                805                 810
```

<210> 51
<211> 5808
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(5805)

<220>
<221> misc_feature
<222> (1)..(5808)
<223> n = a c t or g

<400> 51

```
atg caa ctt cct cca gcg cat tct gcc gat gag aat cgc atc gcg gtc       48
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1               5                   10                  15

gtg ggc atg gcc gtc aaa tat gcg ggc tgt gac aat aaa gaa gag ttt       96
Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
            20                  25                  30

tgg aag act ttg atg aat ggt agt atc aat acc aag tcg att tcg gca      144
Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
        35                  40                  45

gca agg ttg ggc agc aat aag cgt gac gaa cac tat gtt cct gaa cga      192
Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
    50                  55                  60

tcg aaa tat gca gat acg ttc tgt aac gaa agg tac ggt tgt atc cag      240
Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65                  70                  75                  80

caa ggt acg gat aat gag cat gac ctc ctc cta ggt ctt gct caa gaa      288
Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
                85                  90                  95

gct ctc gct gac gct gcc ggg cgg atg gag aaa caa cct tcg gag gcg      336
Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
            100                 105                 110

ttc gat ctg gaa aat act ggc atc gtg agt ggg tgc tta tct ttt cca      384
Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
        115                 120                 125

atg gat aac ctg caa gga gag ttg ttg aac ttg tat caa agc cat gtg      432
Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
    130                 135                 140

gag aaa caa ctt cca cct agt gcc ttg gta gaa gcc gtg aag ctt tgg      480
Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145                 150                 155                 160

tct gag cga cag aaa tct acg aaa gca cat gca ggg gac aag cgc cgg      528
Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
                165                 170                 175
```

```
ttc att gac cca gct tct ttt gta gct gat aaa ctg aac cta ggc cca      576
Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
            180                 185                 190

cta cat tat gcg atc gat gca gca tgc gct tct gca ttg tac gtg tta      624
Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
        195                 200                 205

aaa tta gct caa gac cac ctt gtt tca ggt gcc gtt gat atg atg tta      672
Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
    210                 215                 220

tgt gga gcg acg tgc ttc cca gaa cca ttc ttc atc ttg tct ggg ttc      720
Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225                 230                 235                 240

tcg act ttt caa gcg atg cct gnt ggg gca gat gga gtc tca cta cct      768
Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
                245                 250                 255

ctc cat aaa acg agt gct ggg ctc act cca ggt gaa ggg ggg tcc att      816
Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
            260                 265                 270

atg gtg ctc aag cga ctg aaa gac gct atc aga gat gga aat cac att      864
Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
        275                 280                 285

tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt      912
Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
    290                 295                 300

cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc      960
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305                 310                 315                 320

tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag     1008
Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
                325                 330                 335

tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg     1056
Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
            340                 345                 350

gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa     1104
Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
        355                 360                 365

gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca     1152
Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
    370                 375                 380

aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt     1200
Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385                 390                 395                 400

ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc     1248
Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
                405                 410                 415

act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca     1296
Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
            420                 425                 430

ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat     1344
Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
        435                 440                 445
```

```
gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat     1392
Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
    450                 455                 460

acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc     1440
Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465                 470                 475                 480

ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac     1488
Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
                485                 490                 495

gca gct cgt gac tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac     1536
Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp
            500                 505                 510

ttg gag ttt ctc cga gcc atc agg ctc aag gaa aag cct agg ggt tgt     1584
Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys
        515                 520                 525

ttt gtg gag agt gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg     1632
Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu
    530                 535                 540

aca cca gaa gat atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg     1680
Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr
545                 550                 555                 560

atg gac cga gca att atc gat gca ggt cta aag aag ggc caa cat gta     1728
Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val
                565                 570                 575

gca gtt ctt gtt ggc cta gga act gac ctg gaa ctt tac cgt cat cga     1776
Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg
            580                 585                 590

gca aga gtc gcg ctt aaa gag gtt ttg cac ccg agc tta aag tca gac     1824
Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp
        595                 600                 605

act gca att ctc cag aaa ata atg caa tat gtg aat gat gca gga act     1872
Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr
    610                 615                 620

tcg act tca tac aca tct tac att gga aac ctc gtt gcc acg cgt att     1920
Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile
625                 630                 635                 640

tcg tct cag tgg gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga     1968
Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly
                645                 650                 655

aat aat tcc gtg tac aga tgt gca caa cta gcc aaa gat atg ctt cag     2016
Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln
            660                 665                 670

gtt aac cga gtt gat gct gtc gtc atc gca ggc gtt gat ctc aac gga     2064
Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly
        675                 680                 685

agc gcc gaa agt ttt ttt gtc cga gca aat cgt caa aag ata tcc aag     2112
Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys
    690                 695                 700

cta agt cat cca tgt gca agc ttc gac aga gat gca gat gga ttt ttc     2160
Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe
705                 710                 715                 720
```

```
gca ggt gag ggc tgt ggt gcc cta gtt ttc aag agg tta gaa gac tgt     2208
Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys
                725                 730                 735

gct cct cag gaa aaa att tat gct agt ata gac tct atc gca ata gat     2256
Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp
            740                 745                 750

aaa gag cct act agc tca gct gtg aaa gct gtc tac caa agt gat tcg     2304
Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser
        755                 760                 765

agt ctc tcc gat att gag ctg tta gaa atc agt gga gac tcc aaa cgg     2352
Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg
    770                 775                 780

ttt gca gca ttc gaa ggc gct gtg gaa att caa tca agt gtg gaa gcc     2400
Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala
785                 790                 795                 800

cag cta aaa gga ctt tcc aaa gtc ctt gaa cct gca aaa ggc caa ggc     2448
Gln Leu Lys Gly Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly
                805                 810                 815

gta gcg gtg gga agt act cga gca acc gtt ggg gat ata ggg tat gct     2496
Val Ala Val Gly Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala
            820                 825                 830

aca gga gcg gca agc ctg att aaa act gca ctc tgc tta tat aat cgc     2544
Thr Gly Ala Ala Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg
        835                 840                 845

tac ctt ccg gca tta gca aac tgg agt ggc cca tgt gaa cag tcc gcc     2592
Tyr Leu Pro Ala Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala
    850                 855                 860

tgg ggc tca aac atg ttc gtt tgc cat gaa aca cgg ccg tgg atg aaa     2640
Trp Gly Ser Asn Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys
865                 870                 875                 880

aac cag aat gaa aag aga tgt gcc ctc att tct gga aca gat cca tct     2688
Asn Gln Asn Glu Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser
                885                 890                 895

cat aca tgc ttt tcc ctc gta cta tcg gat act ggg tgt tat gaa gag     2736
His Thr Cys Phe Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu
            900                 905                 910

cac aat cga acg tgc ttt gat gtg caa gcg cca cag cta gtt ctg ata     2784
His Asn Arg Thr Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile
        915                 920                 925

cac gga ttc gat gga aaa act att gtg cgg cga ctt gaa gga tat ctc     2832
His Gly Phe Asp Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu
    930                 935                 940

ctt gaa ctt gtt gaa ggg cat gca agc cct tca gag tat ttc cac aaa     2880
Leu Glu Leu Val Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys
945                 950                 955                 960

ctg att gga caa agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca     2928
Leu Ile Gly Gln Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr
                965                 970                 975

ctt tcg ctt gtg tgc aat ccg aac cag ctc caa aag gag ctc atg ctt     2976
Leu Ser Leu Val Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu
            980                 985                 990
```

```
gct atc aaa gga gta caa cga agc  atg tta aca ggg aag  gat tgg gtc   3024
Ala Ile Lys Gly Val Gln Arg Ser  Met Leu Thr Gly Lys  Asp Trp Val
        995                 1000                1005

agt cca  tca gga agt tgt ttt  gcc cca aat ccg tta  tca agc gca      3069
Ser Pro  Ser Gly Ser Cys Phe  Ala Pro Asn Pro Leu  Ser Ser Ala
    1010                1015                1020

aaa gtg  gca ttc atg tac gga  gaa ggc cga agc ccg  tac tgt ggt      3114
Lys Val  Ala Phe Met Tyr Gly  Glu Gly Arg Ser Pro  Tyr Cys Gly
    1025                1030                1035

gta ggc  ttg ggt cta cat cgt  ttg tgg ccc ggt ctc  cat gaa aat      3159
Val Gly  Leu Gly Leu His Arg  Leu Trp Pro Gly Leu  His Glu Asn
    1040                1045                1050

gtg aac  aat aag aca gtc gat  tta tgg acg gaa gga  gat ggt tgg      3204
Val Asn  Asn Lys Thr Val Asp  Leu Trp Thr Glu Gly  Asp Gly Trp
    1055                1060                1065

tta tat  cct cga acg ttg aca  cga gaa gag cat aca  aaa gcc atc      3249
Leu Tyr  Pro Arg Thr Leu Thr  Arg Glu Glu His Thr  Lys Ala Ile
    1070                1075                1080

gaa tct  ttc aac gca aat caa  att gaa atg ttt cgc  gct ggg att      3294
Glu Ser  Phe Asn Ala Asn Gln  Ile Glu Met Phe Arg  Ala Gly Ile
    1085                1090                1095

ttc atc  tca atg tgt cag aca  gac tat gtc atg aat  gtt ctc ggt      3339
Phe Ile  Ser Met Cys Gln Thr  Asp Tyr Val Met Asn  Val Leu Gly
    1100                1105                1110

gtc cag  cct aag gcc gga ttt  ggg ctg agc ttg gga  gaa att tca      3384
Val Gln  Pro Lys Ala Gly Phe  Gly Leu Ser Leu Gly  Glu Ile Ser
    1115                1120                1125

atg ctc  ttt gcg atg tca aag  gag aac tgc agg cag  tca cag gaa      3429
Met Leu  Phe Ala Met Ser Lys  Glu Asn Cys Arg Gln  Ser Gln Glu
    1130                1135                1140

atg acc  aat cgt ttg cgc ggt  tct cca gtg tgg tct  aac gag ctt      3474
Met Thr  Asn Arg Leu Arg Gly  Ser Pro Val Trp Ser  Asn Glu Leu
    1145                1150                1155

gct atc  aac ttc aat gca att  cgc aag tta tgg aaa  atc ccc cga      3519
Ala Ile  Asn Phe Asn Ala Ile  Arg Lys Leu Trp Lys  Ile Pro Arg
    1160                1165                1170

gga gct  ccc tta gaa tcc ttt  tgg caa gga tac ttg  gtt cac ggc      3564
Gly Ala  Pro Leu Glu Ser Phe  Trp Gln Gly Tyr Leu  Val His Gly
    1175                1180                1185

aca aga  gaa gaa gta gag cat  gct att ggt ctt tct  gag cct tat      3609
Thr Arg  Glu Glu Val Glu His  Ala Ile Gly Leu Ser  Glu Pro Tyr
    1190                1195                1200

gta cgt  ctg ctt att gtg aac  gat tca agg agt gcc  ttg att gct      3654
Val Arg  Leu Leu Ile Val Asn  Asp Ser Arg Ser Ala  Leu Ile Ala
    1205                1210                1215

gga aaa  cca gac gcc tgt cag  gca gta atc agt aga  cta aac tcc      3699
Gly Lys  Pro Asp Ala Cys Gln  Ala Val Ile Ser Arg  Leu Asn Ser
    1220                1225                1230

aag ttc  cct tct ctg ccg gta  aag caa gga atg att  ggt cat tgc      3744
Lys Phe  Pro Ser Leu Pro Val  Lys Gln Gly Met Ile  Gly His Cys
    1235                1240                1245
```

```
cca gaa  gtt cgt gcg ttc atc  aaa gat att ggg tac  atc cat gaa      3789
Pro Glu  Val Arg Ala Phe Ile  Lys Asp Ile Gly Tyr  Ile His Glu
    1250                 1255                 1260

aca ctc  cga att tcc aat gac  tat tcg gat tgt cag  ctt ttc tca      3834
Thr Leu  Arg Ile Ser Asn Asp  Tyr Ser Asp Cys Gln  Leu Phe Ser
    1265                 1270                 1275

gcg gta  acc aag ggc gca ctt  gac agc tcc aca atg  gaa atc aaa      3879
Ala Val  Thr Lys Gly Ala Leu  Asp Ser Ser Thr Met  Glu Ile Lys
    1280                 1285                 1290

cac ttt  gtg gga gag gtc tac  tcc cgg atc gca gac  ttt cct caa      3924
His Phe  Val Gly Glu Val Tyr  Ser Arg Ile Ala Asp  Phe Pro Gln
    1295                 1300                 1305

atc gtc  aac acg gtg cat tcg  gct ggt tat gac gta  ttt ctt gag      3969
Ile Val  Asn Thr Val His Ser  Ala Gly Tyr Asp Val  Phe Leu Glu
    1310                 1315                 1320

ctt ggc  tgt gat gct tct aga  tct gca gca gtt caa  aac att ctt      4014
Leu Gly  Cys Asp Ala Ser Arg  Ser Ala Ala Val Gln  Asn Ile Leu
    1325                 1330                 1335

ggt ggt  caa gga aag ttc ttg  tct aca gct att gac  aaa aaa gga      4059
Gly Gly  Gln Gly Lys Phe Leu  Ser Thr Ala Ile Asp  Lys Lys Gly
    1340                 1345                 1350

cac tcc  gcc tgg tca caa gta  ctt cgg gct acc gca  tca tta gct      4104
His Ser  Ala Trp Ser Gln Val  Leu Arg Ala Thr Ala  Ser Leu Ala
    1355                 1360                 1365

gca cat  cga gta ccg gga atc  tca att ttg gat ttg  ttt cac cca      4149
Ala His  Arg Val Pro Gly Ile  Ser Ile Leu Asp Leu  Phe His Pro
    1370                 1375                 1380

aat ttc  cga gaa atg tgc tgt  aca atg gca acc aca  cct aaa gtg      4194
Asn Phe  Arg Glu Met Cys Cys  Thr Met Ala Thr Thr  Pro Lys Val
    1385                 1390                 1395

gaa gat  aag ttc ctg cgc acg  att caa atc aat ggt  cgg ttt gaa      4239
Glu Asp  Lys Phe Leu Arg Thr  Ile Gln Ile Asn Gly  Arg Phe Glu
    1400                 1405                 1410

aaa gaa  atg att cac cta gaa  gat aca aca tta agt  tgc tta ccc      4284
Lys Glu  Met Ile His Leu Glu  Asp Thr Thr Leu Ser  Cys Leu Pro
    1415                 1420                 1425

gct cca  agt gaa gca aat atc  gca gct att caa tct  cgg tca att      4329
Ala Pro  Ser Glu Ala Asn Ile  Ala Ala Ile Gln Ser  Arg Ser Ile
    1430                 1435                 1440

cga tct  gct gcg gcg cgt tct  gga caa tcc cat gat  tgt gca tcc      4374
Arg Ser  Ala Ala Ala Arg Ser  Gly Gln Ser His Asp  Cys Ala Ser
    1445                 1450                 1455

cat agc  cat gaa gaa aat aag  gat tca tgc cct gaa  aag ctg aag      4419
His Ser  His Glu Glu Asn Lys  Asp Ser Cys Pro Glu  Lys Leu Lys
    1460                 1465                 1470

ctt gat  tct gtg tcc gtc gcc  ata aat ttc gac aat  gat gac cgc      4464
Leu Asp  Ser Val Ser Val Ala  Ile Asn Phe Asp Asn  Asp Asp Arg
    1475                 1480                 1485

att cag  ctt ggg cac gcg ggt  ttt cgg gag atg tac  aat aca aga      4509
Ile Gln  Leu Gly His Ala Gly  Phe Arg Glu Met Tyr  Asn Thr Arg
    1490                 1495                 1500
```

```
tat agc  ttg tac aca ggg gcg  atg gca aag gga att  gca tct gca      4554
Tyr Ser  Leu Tyr Thr Gly Ala  Met Ala Lys Gly Ile  Ala Ser Ala
    1505                 1510                 1515

gat ctt  gtc att gcc gct ggg  aaa gag ggc atc cta  gct tcc tat      4599
Asp Leu  Val Ile Ala Ala Gly  Lys Glu Gly Ile Leu  Ala Ser Tyr
    1520                 1525                 1530

gga gct  gga gga cta cct ctt  gct act gtt cga aag  gga ata gac      4644
Gly Ala  Gly Gly Leu Pro Leu  Ala Thr Val Arg Lys  Gly Ile Asp
    1535                 1540                 1545

aaa att  caa caa gcc ttg cca  agt ggc cca tat gct  gta aat ctt      4689
Lys Ile  Gln Gln Ala Leu Pro  Ser Gly Pro Tyr Ala  Val Asn Leu
    1550                 1555                 1560

att cac  tct ccc ttt gac ggc  aac ttg gag cag gga  aac gtc gat      4734
Ile His  Ser Pro Phe Asp Gly  Asn Leu Glu Gln Gly  Asn Val Asp
    1565                 1570                 1575

ttg ttc  ttg gaa aag aac gtc  cgc gtg gcg gaa tgt  tcc gcg ttt      4779
Leu Phe  Leu Glu Lys Asn Val  Arg Val Ala Glu Cys  Ser Ala Phe
    1580                 1585                 1590

aca acg  cta aca gtg cca gta  gta cac tat cgt gct  gca ggg ctt      4824
Thr Thr  Leu Thr Val Pro Val  Val His Tyr Arg Ala  Ala Gly Leu
    1595                 1600                 1605

gtt cgg  cgc caa gat gga agc  att ttg atc aag aac  cga atc att      4869
Val Arg  Arg Gln Asp Gly Ser  Ile Leu Ile Lys Asn  Arg Ile Ile
    1610                 1615                 1620

gct aaa  gta tct agg aca gaa  ctc gct gag atg ttc  ctt cgt ccg      4914
Ala Lys  Val Ser Arg Thr Glu  Leu Ala Glu Met Phe  Leu Arg Pro
    1625                 1630                 1635

gca cct  caa atc atc ctc gaa  aaa ctg gta gca gca  gaa atc att      4959
Ala Pro  Gln Ile Ile Leu Glu  Lys Leu Val Ala Ala  Glu Ile Ile
    1640                 1645                 1650

tca tct  gac caa gcg cgt atg  gca gcc aaa gtt ccc  atg gcg gac      5004
Ser Ser  Asp Gln Ala Arg Met  Ala Ala Lys Val Pro  Met Ala Asp
    1655                 1660                 1665

gac atc  gca gtc gaa gcc gac  tct ggt ggg cac acg  gat aat cgg      5049
Asp Ile  Ala Val Glu Ala Asp  Ser Gly Gly His Thr  Asp Asn Arg
    1670                 1675                 1680

cct atg  cac gtc att ttg ccc  ctg ata att caa ctc  cgc aat act      5094
Pro Met  His Val Ile Leu Pro  Leu Ile Ile Gln Leu  Arg Asn Thr
    1685                 1690                 1695

ata ctt  gca gag tat ggc tgt  gcc acg gct ttt cgt  acc cgt ata      5139
Ile Leu  Ala Glu Tyr Gly Cys  Ala Thr Ala Phe Arg  Thr Arg Ile
    1700                 1705                 1710

ggc gct  gga gga ggc att ggt  tgt cct tca gcg gcc  ctc gca gcc      5184
Gly Ala  Gly Gly Gly Ile Gly  Cys Pro Ser Ala Ala  Leu Ala Ala
    1715                 1720                 1725

ttt gat  atg ggt gcg agt ttt  gtc gtg act gga agc  ata aat caa      5229
Phe Asp  Met Gly Ala Ser Phe  Val Val Thr Gly Ser  Ile Asn Gln
    1730                 1735                 1740

att tgc  cgc gag gca ggg act  tgc gat act gtt cgg  gag cta ctt      5274
Ile Cys  Arg Glu Ala Gly Thr  Cys Asp Thr Val Arg  Glu Leu Leu
    1745                 1750                 1755
```

```
gcc aac  tca agc tac tcg gac  gtg acg atg gcg cca  gca gca gac      5319
Ala Asn  Ser Ser Tyr Ser Asp  Val Thr Met Ala Pro  Ala Ala Asp
    1760                 1765                 1770

atg ttt  gac caa ggt gtg aaa  ctc caa gtc tta aaa  cga gga acg      5364
Met Phe  Asp Gln Gly Val Lys  Leu Gln Val Leu Lys  Arg Gly Thr
    1775                 1780                 1785

atg ttt  cca agc aga gca aat  aaa ctc cgg aag ctc  ttt gtg aac      5409
Met Phe  Pro Ser Arg Ala Asn  Lys Leu Arg Lys Leu  Phe Val Asn
    1790                 1795                 1800

tac gaa  tct cta gaa aca ctc  ccg tcg aaa gag ttg  aaa tac ctg      5454
Tyr Glu  Ser Leu Glu Thr Leu  Pro Ser Lys Glu Leu  Lys Tyr Leu
    1805                 1810                 1815

gaa aac  atc ata ttc aag caa  gca gta gac cag gtg  tgg gag gaa      5499
Glu Asn  Ile Ile Phe Lys Gln  Ala Val Asp Gln Val  Trp Glu Glu
    1820                 1825                 1830

aca aag  cgc ttt tac tgt gaa  aaa ctg aac aat cca  gat aaa att      5544
Thr Lys  Arg Phe Tyr Cys Glu  Lys Leu Asn Asn Pro  Asp Lys Ile
    1835                 1840                 1845

gca agg  gcc atg aaa gat cct  aaa ttg aag atg tcg  ctt tgc ttt      5589
Ala Arg  Ala Met Lys Asp Pro  Lys Leu Lys Met Ser  Leu Cys Phe
    1850                 1855                 1860

cgg tgg  tat ctc tcc aag agc  tct ggg tgg gcc aac  gca gga att      5634
Arg Trp  Tyr Leu Ser Lys Ser  Ser Gly Trp Ala Asn  Ala Gly Ile
    1865                 1870                 1875

aaa tct  cgt gca ctc gac tac  cag atc tgg tgt ggc  ccg gca atg      5679
Lys Ser  Arg Ala Leu Asp Tyr  Gln Ile Trp Cys Gly  Pro Ala Met
    1880                 1885                 1890

ggc tcg  ttc aac aat ttc gcc  agc ggc aca tcc ctc  gat tgg aaa      5724
Gly Ser  Phe Asn Asn Phe Ala  Ser Gly Thr Ser Leu  Asp Trp Lys
    1895                 1900                 1905

gtg act  ggg gtt ttc cct ggc  gtt gcg gaa gta aac  atg gcc att      5769
Val Thr  Gly Val Phe Pro Gly  Val Ala Glu Val Asn  Met Ala Ile
    1910                 1915                 1920

tta gat  ggc gcg cga gaa cta  gct gct aaa cga aat  taa              5808
Leu Asp  Gly Ala Arg Glu Leu  Ala Ala Lys Arg Asn
    1925                 1930                 1935
```

<210> 52
<211> 1935
<212> PRT
<213> Thraustochytrium sp.

<220>
<221> misc_feature
<222> (248)..(248)
<223> The 'Xaa' at location 248 stands for Asp, Gly, Ala, or Val.

<400> 52

```
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1               5                  10                  15


Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
            20                  25                  30
```

```
Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
        35                  40                  45

Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
    50                  55                  60

Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65                  70                  75                  80

Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
                85                  90                  95

Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
            100                 105                 110

Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
        115                 120                 125

Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
    130                 135                 140

Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145                 150                 155                 160

Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
                165                 170                 175

Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
            180                 185                 190

Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
        195                 200                 205

Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
    210                 215                 220

Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225                 230                 235                 240

Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
                245                 250                 255

Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
            260                 265                 270

Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
        275                 280                 285

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
    290                 295                 300
```

```
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305                 310                 315                 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
                325                 330                 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
            340                 345                 350

Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
        355                 360                 365

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
    370                 375                 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385                 390                 395                 400

Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
                405                 410                 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
            420                 425                 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
        435                 440                 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
    450                 455                 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465                 470                 475                 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
                485                 490                 495

Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp
            500                 505                 510

Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys
        515                 520                 525

Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu
    530                 535                 540

Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr
545                 550                 555                 560

Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val
                565                 570                 575
```

```
Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg
            580                 585                 590

Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp
        595                 600                 605

Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr
    610                 615                 620

Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile
625                 630                 635                 640

Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly
                645                 650                 655

Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln
            660                 665                 670

Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly
        675                 680                 685

Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys
    690                 695                 700

Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe
705                 710                 715                 720

Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys
                725                 730                 735

Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp
            740                 745                 750

Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser
        755                 760                 765

Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg
    770                 775                 780

Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala
785                 790                 795                 800

Gln Leu Lys Gly Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly
                805                 810                 815

Val Ala Val Gly Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala
            820                 825                 830

Thr Gly Ala Ala Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg
        835                 840                 845
```

```
Tyr Leu Pro Ala Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala
    850                 855                 860

Trp Gly Ser Asn Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys
865                 870                 875                 880

Asn Gln Asn Glu Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser
                885                 890                 895

His Thr Cys Phe Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu
            900                 905                 910

His Asn Arg Thr Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile
        915                 920                 925

His Gly Phe Asp Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu
    930                 935                 940

Leu Glu Leu Val Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys
945                 950                 955                 960

Leu Ile Gly Gln Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr
                965                 970                 975

Leu Ser Leu Val Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu
            980                 985                 990

Ala Ile Lys Gly Val Gln Arg Ser  Met Leu Thr Gly Lys  Asp Trp Val
        995                 1000                1005

Ser Pro  Ser Gly Ser Cys Phe  Ala Pro Asn Pro Leu  Ser Ser Ala
    1010                1015                1020

Lys Val  Ala Phe Met Tyr Gly  Glu Gly Arg Ser Pro  Tyr Cys Gly
    1025                1030                1035

Val Gly  Leu Gly Leu His Arg  Leu Trp Pro Gly Leu  His Glu Asn
    1040                1045                1050

Val Asn  Asn Lys Thr Val Asp  Leu Trp Thr Glu Gly  Asp Gly Trp
    1055                1060                1065

Leu Tyr  Pro Arg Thr Leu Thr  Arg Glu Glu His Thr  Lys Ala Ile
    1070                1075                1080

Glu Ser  Phe Asn Ala Asn Gln  Ile Glu Met Phe Arg  Ala Gly Ile
    1085                1090                1095

Phe Ile  Ser Met Cys Gln Thr  Asp Tyr Val Met Asn  Val Leu Gly
    1100                1105                1110
```

```
Val Gln  Pro Lys Ala Gly Phe  Gly Leu Ser Leu Gly  Glu Ile Ser
    1115                1120                1125

Met Leu  Phe Ala Met Ser Lys  Glu Asn Cys Arg Gln  Ser Gln Glu
    1130                1135                1140

Met Thr  Asn Arg Leu Arg Gly  Ser Pro Val Trp Ser  Asn Glu Leu
    1145                1150                1155

Ala Ile  Asn Phe Asn Ala Ile  Arg Lys Leu Trp Lys  Ile Pro Arg
    1160                1165                1170

Gly Ala  Pro Leu Glu Ser Phe  Trp Gln Gly Tyr Leu  Val His Gly
    1175                1180                1185

Thr Arg  Glu Glu Val Glu His  Ala Ile Gly Leu Ser  Glu Pro Tyr
    1190                1195                1200

Val Arg  Leu Leu Ile Val Asn  Asp Ser Arg Ser Ala  Leu Ile Ala
    1205                1210                1215

Gly Lys  Pro Asp Ala Cys Gln  Ala Val Ile Ser Arg  Leu Asn Ser
    1220                1225                1230

Lys Phe  Pro Ser Leu Pro Val  Lys Gln Gly Met Ile  Gly His Cys
    1235                1240                1245

Pro Glu  Val Arg Ala Phe Ile  Lys Asp Ile Gly Tyr  Ile His Glu
    1250                1255                1260

Thr Leu  Arg Ile Ser Asn Asp  Tyr Ser Asp Cys Gln  Leu Phe Ser
    1265                1270                1275

Ala Val  Thr Lys Gly Ala Leu  Asp Ser Ser Thr Met  Glu Ile Lys
    1280                1285                1290

His Phe  Val Gly Glu Val Tyr  Ser Arg Ile Ala Asp  Phe Pro Gln
    1295                1300                1305

Ile Val  Asn Thr Val His Ser  Ala Gly Tyr Asp Val  Phe Leu Glu
    1310                1315                1320

Leu Gly  Cys Asp Ala Ser Arg  Ser Ala Ala Val Gln  Asn Ile Leu
    1325                1330                1335

Gly Gly  Gln Gly Lys Phe Leu  Ser Thr Ala Ile Asp  Lys Lys Gly
    1340                1345                1350

His Ser  Ala Trp Ser Gln Val  Leu Arg Ala Thr Ala  Ser Leu Ala
    1355                1360                1365
```

```
Ala His  Arg Val Pro Gly Ile  Ser Ile Leu Asp Leu  Phe His Pro
    1370                1375                1380

Asn Phe  Arg Glu Met Cys Cys  Thr Met Ala Thr Thr  Pro Lys Val
    1385                1390                1395

Glu Asp  Lys Phe Leu Arg Thr  Ile Gln Ile Asn Gly  Arg Phe Glu
    1400                1405                1410

Lys Glu  Met Ile His Leu Glu  Asp Thr Thr Leu Ser  Cys Leu Pro
    1415                1420                1425

Ala Pro  Ser Glu Ala Asn Ile  Ala Ala Ile Gln Ser  Arg Ser Ile
    1430                1435                1440

Arg Ser  Ala Ala Ala Arg Ser  Gly Gln Ser His Asp  Cys Ala Ser
    1445                1450                1455

His Ser  His Glu Glu Asn Lys  Asp Ser Cys Pro Glu  Lys Leu Lys
    1460                1465                1470

Leu Asp  Ser Val Ser Val Ala  Ile Asn Phe Asp Asn  Asp Asp Arg
    1475                1480                1485

Ile Gln  Leu Gly His Ala Gly  Phe Arg Glu Met Tyr  Asn Thr Arg
    1490                1495                1500

Tyr Ser  Leu Tyr Thr Gly Ala  Met Ala Lys Gly Ile  Ala Ser Ala
    1505                1510                1515

Asp Leu  Val Ile Ala Ala Gly  Lys Glu Gly Ile Leu  Ala Ser Tyr
    1520                1525                1530

Gly Ala  Gly Gly Leu Pro Leu  Ala Thr Val Arg Lys  Gly Ile Asp
    1535                1540                1545

Lys Ile  Gln Gln Ala Leu Pro  Ser Gly Pro Tyr Ala  Val Asn Leu
    1550                1555                1560

Ile His  Ser Pro Phe Asp Gly  Asn Leu Glu Gln Gly  Asn Val Asp
    1565                1570                1575

Leu Phe  Leu Glu Lys Asn Val  Arg Val Ala Glu Cys  Ser Ala Phe
    1580                1585                1590

Thr Thr  Leu Thr Val Pro Val  Val His Tyr Arg Ala  Ala Gly Leu
    1595                1600                1605

Val Arg  Arg Gln Asp Gly Ser  Ile Leu Ile Lys Asn  Arg Ile Ile
    1610                1615                1620
```

```
Ala Lys  Val Ser Arg Thr Glu  Leu Ala Glu Met Phe  Leu Arg Pro
    1625                1630                1635

Ala Pro  Gln Ile Ile Leu Glu  Lys Leu Val Ala Ala  Glu Ile Ile
    1640                1645                1650

Ser Ser  Asp Gln Ala Arg Met  Ala Ala Lys Val Pro  Met Ala Asp
    1655                1660                1665

Asp Ile  Ala Val Glu Ala Asp  Ser Gly Gly His Thr  Asp Asn Arg
    1670                1675                1680

Pro Met  His Val Ile Leu Pro  Leu Ile Ile Gln Leu  Arg Asn Thr
    1685                1690                1695

Ile Leu  Ala Glu Tyr Gly Cys  Ala Thr Ala Phe Arg  Thr Arg Ile
    1700                1705                1710

Gly Ala  Gly Gly Gly Ile Gly  Cys Pro Ser Ala Ala  Leu Ala Ala
    1715                1720                1725

Phe Asp  Met Gly Ala Ser Phe  Val Val Thr Gly Ser  Ile Asn Gln
    1730                1735                1740

Ile Cys  Arg Glu Ala Gly Thr  Cys Asp Thr Val Arg  Glu Leu Leu
    1745                1750                1755

Ala Asn  Ser Ser Tyr Ser Asp  Val Thr Met Ala Pro  Ala Ala Asp
    1760                1765                1770

Met Phe  Asp Gln Gly Val Lys  Leu Gln Val Leu Lys  Arg Gly Thr
    1775                1780                1785

Met Phe  Pro Ser Arg Ala Asn  Lys Leu Arg Lys Leu  Phe Val Asn
    1790                1795                1800

Tyr Glu  Ser Leu Glu Thr Leu  Pro Ser Lys Glu Leu  Lys Tyr Leu
    1805                1810                1815

Glu Asn  Ile Ile Phe Lys Gln  Ala Val Asp Gln Val  Trp Glu Glu
    1820                1825                1830

Thr Lys  Arg Phe Tyr Cys Glu  Lys Leu Asn Asn Pro  Asp Lys Ile
    1835                1840                1845

Ala Arg  Ala Met Lys Asp Pro  Lys Leu Lys Met Ser  Leu Cys Phe
    1850                1855                1860

Arg Trp  Tyr Leu Ser Lys Ser  Ser Gly Trp Ala Asn  Ala Gly Ile
    1865                1870                1875
```

```
Lys Ser  Arg Ala Leu Asp Tyr  Gln Ile Trp Cys Gly  Pro Ala Met
    1880                 1885                 1890


Gly Ser  Phe Asn Asn Phe Ala  Ser Gly Thr Ser Leu  Asp Trp Lys
    1895                 1900                 1905


Val Thr  Gly Val Phe Pro Gly  Val Ala Glu Val Asn  Met Ala Ile
    1910                 1915                 1920


Leu Asp  Gly Ala Arg Glu Leu  Ala Ala Lys Arg Asn
    1925                 1930                 1935


<210>  53
<211>  1500
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<220>
<221>  misc_feature
<222>  (1)..(1500)
<223>  n = a c t or g

<400>  53
atg caa ctt cct cca gcg cat tct gcc gat gag aat cgc atc gcg gtc       48
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1               5                   10                  15

gtg ggc atg gcc gtc aaa tat gcg ggc tgt gac aat aaa gaa gag ttt       96
Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
            20                  25                  30

tgg aag act ttg atg aat ggt agt atc aat acc aag tcg att tcg gca      144
Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
        35                  40                  45

gca agg ttg ggc agc aat aag cgt gac gaa cac tat gtt cct gaa cga      192
Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
    50                  55                  60

tcg aaa tat gca gat acg ttc tgt aac gaa agg tac ggt tgt atc cag      240
Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65                  70                  75                  80

caa ggt acg gat aat gag cat gac ctc ctc cta ggt ctt gct caa gaa      288
Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
                85                  90                  95

gct ctc gct gac gct gcc ggg cgg atg gag aaa caa cct tcg gag gcg      336
Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
            100                 105                 110

ttc gat ctg gaa aat act ggc atc gtg agt ggg tgc tta tct ttt cca      384
Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
        115                 120                 125

atg gat aac ctg caa gga gag ttg ttg aac ttg tat caa agc cat gtg      432
Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
    130                 135                 140
```

```
gag aaa caa ctt cca cct agt gcc ttg gta gaa gcc gtg aag ctt tgg      480
Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145                 150                 155                 160

tct gag cga cag aaa tct acg aaa gca cat gca ggg gac aag cgc cgg      528
Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
                165                 170                 175

ttc att gac cca gct tct ttt gta gct gat aaa ctg aac cta ggc cca      576
Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
            180                 185                 190

cta cat tat gcg atc gat gca gca tgc gct tct gca ttg tac gtg tta      624
Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
        195                 200                 205

aaa tta gct caa gac cac ctt gtt tca ggt gcc gtt gat atg atg tta      672
Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
    210                 215                 220

tgt gga gcg acg tgc ttc cca gaa cca ttc ttc atc ttg tct ggg ttc      720
Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225                 230                 235                 240

tcg act ttt caa gcg atg cct gnt ggg gca gat gga gtc tca cta cct      768
Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
                245                 250                 255

ctc cat aaa acg agt gct ggg ctc act cca ggt gaa ggg ggg tcc att      816
Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
            260                 265                 270

atg gtg ctc aag cga ctg aaa gac gct atc aga gat gga aat cac att      864
Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
        275                 280                 285

tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt      912
Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
    290                 295                 300

cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc      960
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305                 310                 315                 320

tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag     1008
Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
                325                 330                 335

tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg     1056
Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
            340                 345                 350

gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa     1104
Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
        355                 360                 365

gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca     1152
Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
    370                 375                 380

aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt     1200
Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385                 390                 395                 400

ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc     1248
Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
                405                 410                 415
```

```
act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca     1296
Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
            420                 425                 430

ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat     1344
Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
        435                 440                 445

gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat     1392
Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
    450                 455                 460

acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc     1440
Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465                 470                 475                 480

ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac     1488
Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
                485                 490                 495

gca gct cgt gac                                                    1500
Ala Ala Arg Asp
            500
```

<210> 54
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<220>
<221> misc_feature
<222> (248)..(248)
<223> The 'Xaa' at location 248 stands for Asp, Gly, Ala, or Val.

<400> 54

```
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1               5                   10                  15

Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
            20                  25                  30

Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
        35                  40                  45

Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
    50                  55                  60

Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65                  70                  75                  80

Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
                85                  90                  95

Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
            100                 105                 110

Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
        115                 120                 125
```

```
Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
    130                 135                 140

Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145                 150                 155                 160

Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
                165                 170                 175

Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
            180                 185                 190

Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
        195                 200                 205

Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
    210                 215                 220

Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225                 230                 235                 240

Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
                245                 250                 255

Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
            260                 265                 270

Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
        275                 280                 285

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
    290                 295                 300

Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305                 310                 315                 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
                325                 330                 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
            340                 345                 350

Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
        355                 360                 365

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
    370                 375                 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385                 390                 395                 400
```

```
Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
                405                 410                 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
            420                 425                 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
        435                 440                 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
    450                 455                 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465                 470                 475                 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
                485                 490                 495

Ala Ala Arg Asp
            500


<210>  55
<211>  1500
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<400>  55
tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac ttg gag ttt ctc       48
Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp Leu Glu Phe Leu
1               5                   10                  15

cga gcc atc agg ctc aag gaa aag cct agg ggt tgt ttt gtg gag agt       96
Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys Phe Val Glu Ser
            20                  25                  30

gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg aca cca gaa gat      144
Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu Thr Pro Glu Asp
        35                  40                  45

atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg atg gac cga gca      192
Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr Met Asp Arg Ala
    50                  55                  60

att atc gat gca ggt cta aag aag ggc caa cat gta gca gtt ctt gtt      240
Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val
65                  70                  75                  80

ggc cta gga act gac ctg gaa ctt tac cgt cat cga gca aga gtc gcg      288
Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala
                85                  90                  95

ctt aaa gag gtt ttg cac ccg agc tta aag tca gac act gca att ctc      336
Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu
            100                 105                 110
```

```
cag aaa ata atg caa tat gtg aat gat gca gga act tcg act tca tac      384
Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr
        115                 120                 125

aca tct tac att gga aac ctc gtt gcc acg cgt att tcg tct cag tgg      432
Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp
    130                 135                 140

gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga aat aat tcc gtg      480
Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly Asn Asn Ser Val
145                 150                 155                 160

tac aga tgt gca caa cta gcc aaa gat atg ctt cag gtt aac cga gtt      528
Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln Val Asn Arg Val
                165                 170                 175

gat gct gtc gtc atc gca ggc gtt gat ctc aac gga agc gcc gaa agt      576
Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly Ser Ala Glu Ser
            180                 185                 190

ttt ttt gtc cga gca aat cgt caa aag ata tcc aag cta agt cat cca      624
Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys Leu Ser His Pro
        195                 200                 205

tgt gca agc ttc gac aga gat gca gat gga ttt ttc gca ggt gag ggc      672
Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe Ala Gly Glu Gly
    210                 215                 220

tgt ggt gcc cta gtt ttc aag agg tta gaa gac tgt gct cct cag gaa      720
Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys Ala Pro Gln Glu
225                 230                 235                 240

aaa att tat gct agt ata gac tct atc gca ata gat aaa gag cct act      768
Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp Lys Glu Pro Thr
                245                 250                 255

agc tca gct gtg aaa gct gtc tac caa agt gat tcg agt ctc tcc gat      816
Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser Ser Leu Ser Asp
            260                 265                 270

att gag ctg tta gaa atc agt gga gac tcc aaa cgg ttt gca gca ttc      864
Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg Phe Ala Ala Phe
        275                 280                 285

gaa ggc gct gtg gaa att caa tca agt gtg gaa gcc cag cta aaa gga      912
Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala Gln Leu Lys Gly
    290                 295                 300

ctt tcc aaa gtc ctt gaa cct gca aaa ggc caa ggc gta gcg gtg gga      960
Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly Val Ala Val Gly
305                 310                 315                 320

agt act cga gca acc gtt ggg gat ata ggg tat gct aca gga gcg gca     1008
Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala Thr Gly Ala Ala
                325                 330                 335

agc ctg att aaa act gca ctc tgc tta tat aat cgc tac ctt ccg gca     1056
Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg Tyr Leu Pro Ala
            340                 345                 350

tta gca aac tgg agt ggc cca tgt gaa cag tcc gcc tgg ggc tca aac     1104
Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala Trp Gly Ser Asn
        355                 360                 365

atg ttc gtt tgc cat gaa aca cgg ccg tgg atg aaa aac cag aat gaa     1152
Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys Asn Gln Asn Glu
    370                 375                 380
```

```
aag aga tgt gcc ctc att tct gga aca gat cca tct cat aca tgc ttt     1200
Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser His Thr Cys Phe
385                 390                 395                 400

tcc ctc gta cta tcg gat act ggg tgt tat gaa gag cac aat cga acg     1248
Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu His Asn Arg Thr
                405                 410                 415

tgc ttt gat gtg caa gcg cca cag cta gtt ctg ata cac gga ttc gat     1296
Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile His Gly Phe Asp
            420                 425                 430

gga aaa act att gtg cgg cga ctt gaa gga tat ctc ctt gaa ctt gtt     1344
Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val
        435                 440                 445

gaa ggg cat gca agc cct tca gag tat ttc cac aaa ctg att gga caa     1392
Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln
    450                 455                 460

agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca ctt tcg ctt gtg     1440
Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val
465                 470                 475                 480

tgc aat ccg aac cag ctc caa aag gag ctc atg ctt gct atc aaa gga     1488
Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly
                485                 490                 495

gta caa cga agc                                                    1500
Val Gln Arg Ser
            500
```

<210> 56
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 56

```
Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp Leu Glu Phe Leu
1               5                   10                  15

Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys Phe Val Glu Ser
            20                  25                  30

Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu Thr Pro Glu Asp
        35                  40                  45

Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr Met Asp Arg Ala
    50                  55                  60

Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val
65                  70                  75                  80

Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala
                85                  90                  95

Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu
            100                 105                 110
```

```
Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr
        115                 120                 125

Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp
    130                 135                 140

Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly Asn Asn Ser Val
145                 150                 155                 160

Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln Val Asn Arg Val
                165                 170                 175

Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly Ser Ala Glu Ser
            180                 185                 190

Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys Leu Ser His Pro
        195                 200                 205

Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe Ala Gly Glu Gly
    210                 215                 220

Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys Ala Pro Gln Glu
225                 230                 235                 240

Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp Lys Glu Pro Thr
                245                 250                 255

Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser Ser Leu Ser Asp
            260                 265                 270

Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg Phe Ala Ala Phe
        275                 280                 285

Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala Gln Leu Lys Gly
    290                 295                 300

Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly Val Ala Val Gly
305                 310                 315                 320

Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala Thr Gly Ala Ala
                325                 330                 335

Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg Tyr Leu Pro Ala
            340                 345                 350

Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala Trp Gly Ser Asn
        355                 360                 365

Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys Asn Gln Asn Glu
    370                 375                 380
```

```
Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser His Thr Cys Phe
385                 390                 395                 400


Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu His Asn Arg Thr
                405                 410                 415


Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile His Gly Phe Asp
            420                 425                 430


Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val
        435                 440                 445


Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln
    450                 455                 460


Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val
465                 470                 475                 480


Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly
                485                 490                 495


Val Gln Arg Ser
            500


<210>  57
<211>  1500
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<400>  57
atg tta aca ggg aag gat tgg gtc agt cca tca gga agt tgt ttt gcc       48
Met Leu Thr Gly Lys Asp Trp Val Ser Pro Ser Gly Ser Cys Phe Ala
1               5                   10                  15

cca aat ccg tta tca agc gca aaa gtg gca ttc atg tac gga gaa ggc       96
Pro Asn Pro Leu Ser Ser Ala Lys Val Ala Phe Met Tyr Gly Glu Gly
            20                  25                  30

cga agc ccg tac tgt ggt gta ggc ttg ggt cta cat cgt ttg tgg ccc      144
Arg Ser Pro Tyr Cys Gly Val Gly Leu Gly Leu His Arg Leu Trp Pro
        35                  40                  45

ggt ctc cat gaa aat gtg aac aat aag aca gtc gat tta tgg acg gaa      192
Gly Leu His Glu Asn Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu
    50                  55                  60

gga gat ggt tgg tta tat cct cga acg ttg aca cga gaa gag cat aca      240
Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
65                  70                  75                  80

aaa gcc atc gaa tct ttc aac gca aat caa att gaa atg ttt cgc gct      288
Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala
                85                  90                  95
```

```
ggg att ttc atc tca atg tgt cag aca gac tat gtc atg aat gtt ctc      336
Gly Ile Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu
            100                 105                 110

ggt gtc cag cct aag gcc gga ttt ggg ctg agc ttg gga gaa att tca      384
Gly Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser
        115                 120                 125

atg ctc ttt gcg atg tca aag gag aac tgc agg cag tca cag gaa atg      432
Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu Met
    130                 135                 140

acc aat cgt ttg cgc ggt tct cca gtg tgg tct aac gag ctt gct atc      480
Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu Ala Ile
145                 150                 155                 160

aac ttc aat gca att cgc aag tta tgg aaa atc ccc cga gga gct ccc      528
Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg Gly Ala Pro
                165                 170                 175

tta gaa tcc ttt tgg caa gga tac ttg gtt cac ggc aca aga gaa gaa      576
Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly Thr Arg Glu Glu
            180                 185                 190

gta gag cat gct att ggt ctt tct gag cct tat gta cgt ctg ctt att      624
Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr Val Arg Leu Leu Ile
        195                 200                 205

gtg aac gat tca agg agt gcc ttg att gct gga aaa cca gac gcc tgt      672
Val Asn Asp Ser Arg Ser Ala Leu Ile Ala Gly Lys Pro Asp Ala Cys
    210                 215                 220

cag gca gta atc agt aga cta aac tcc aag ttc cct tct ctg ccg gta      720
Gln Ala Val Ile Ser Arg Leu Asn Ser Lys Phe Pro Ser Leu Pro Val
225                 230                 235                 240

aag caa gga atg att ggt cat tgc cca gaa gtt cgt gcg ttc atc aaa      768
Lys Gln Gly Met Ile Gly His Cys Pro Glu Val Arg Ala Phe Ile Lys
                245                 250                 255

gat att ggg tac atc cat gaa aca ctc cga att tcc aat gac tat tcg      816
Asp Ile Gly Tyr Ile His Glu Thr Leu Arg Ile Ser Asn Asp Tyr Ser
            260                 265                 270

gat tgt cag ctt ttc tca gcg gta acc aag ggc gca ctt gac agc tcc      864
Asp Cys Gln Leu Phe Ser Ala Val Thr Lys Gly Ala Leu Asp Ser Ser
        275                 280                 285

aca atg gaa atc aaa cac ttt gtg gga gag gtc tac tcc cgg atc gca      912
Thr Met Glu Ile Lys His Phe Val Gly Glu Val Tyr Ser Arg Ile Ala
    290                 295                 300

gac ttt cct caa atc gtc aac acg gtg cat tcg gct ggt tat gac gta      960
Asp Phe Pro Gln Ile Val Asn Thr Val His Ser Ala Gly Tyr Asp Val
305                 310                 315                 320

ttt ctt gag ctt ggc tgt gat gct tct aga tct gca gca gtt caa aac     1008
Phe Leu Glu Leu Gly Cys Asp Ala Ser Arg Ser Ala Ala Val Gln Asn
                325                 330                 335

att ctt ggt ggt caa gga aag ttc ttg tct aca gct att gac aaa aaa     1056
Ile Leu Gly Gly Gln Gly Lys Phe Leu Ser Thr Ala Ile Asp Lys Lys
            340                 345                 350

gga cac tcc gcc tgg tca caa gta ctt cgg gct acc gca tca tta gct     1104
Gly His Ser Ala Trp Ser Gln Val Leu Arg Ala Thr Ala Ser Leu Ala
        355                 360                 365
```

```
gca cat cga gta ccg gga atc tca att ttg gat ttg ttt cac cca aat     1152
Ala His Arg Val Pro Gly Ile Ser Ile Leu Asp Leu Phe His Pro Asn
    370                 375                 380

ttc cga gaa atg tgc tgt aca atg gca acc aca cct aaa gtg gaa gat     1200
Phe Arg Glu Met Cys Cys Thr Met Ala Thr Thr Pro Lys Val Glu Asp
385                 390                 395                 400

aag ttc ctg cgc acg att caa atc aat ggt cgg ttt gaa aaa gaa atg     1248
Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu Lys Glu Met
                405                 410                 415

att cac cta gaa gat aca aca tta agt tgc tta ccc gct cca agt gaa     1296
Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro Ala Pro Ser Glu
            420                 425                 430

gca aat atc gca gct att caa tct cgg tca att cga tct gct gcg gcg     1344
Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile Arg Ser Ala Ala Ala
        435                 440                 445

cgt tct gga caa tcc cat gat tgt gca tcc cat agc cat gaa gaa aat     1392
Arg Ser Gly Gln Ser His Asp Cys Ala Ser His Ser His Glu Glu Asn
    450                 455                 460

aag gat tca tgc cct gaa aag ctg aag ctt gat tct gtg tcc gtc gcc     1440
Lys Asp Ser Cys Pro Glu Lys Leu Lys Leu Asp Ser Val Ser Val Ala
465                 470                 475                 480

ata aat ttc gac aat gat gac cgc att cag ctt ggg cac gcg ggt ttt     1488
Ile Asn Phe Asp Asn Asp Asp Arg Ile Gln Leu Gly His Ala Gly Phe
                485                 490                 495

cgg gag atg tac                                                    1500
Arg Glu Met Tyr
            500
```

<210> 58
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 58

```
Met Leu Thr Gly Lys Asp Trp Val Ser Pro Ser Gly Ser Cys Phe Ala
1               5                   10                  15

Pro Asn Pro Leu Ser Ser Ala Lys Val Ala Phe Met Tyr Gly Glu Gly
            20                  25                  30

Arg Ser Pro Tyr Cys Gly Val Gly Leu Gly Leu His Arg Leu Trp Pro
        35                  40                  45

Gly Leu His Glu Asn Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu
    50                  55                  60

Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
65                  70                  75                  80

Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala
                85                  90                  95
```

```
Gly Ile Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu
            100                 105                 110

Gly Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser
        115                 120                 125

Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu Met
    130                 135                 140

Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu Ala Ile
145                 150                 155                 160

Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg Gly Ala Pro
                165                 170                 175

Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly Thr Arg Glu Glu
            180                 185                 190

Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr Val Arg Leu Leu Ile
        195                 200                 205

Val Asn Asp Ser Arg Ser Ala Leu Ile Ala Gly Lys Pro Asp Ala Cys
    210                 215                 220

Gln Ala Val Ile Ser Arg Leu Asn Ser Lys Phe Pro Ser Leu Pro Val
225                 230                 235                 240

Lys Gln Gly Met Ile Gly His Cys Pro Glu Val Arg Ala Phe Ile Lys
                245                 250                 255

Asp Ile Gly Tyr Ile His Glu Thr Leu Arg Ile Ser Asn Asp Tyr Ser
            260                 265                 270

Asp Cys Gln Leu Phe Ser Ala Val Thr Lys Gly Ala Leu Asp Ser Ser
        275                 280                 285

Thr Met Glu Ile Lys His Phe Val Gly Glu Val Tyr Ser Arg Ile Ala
    290                 295                 300

Asp Phe Pro Gln Ile Val Asn Thr Val His Ser Ala Gly Tyr Asp Val
305                 310                 315                 320

Phe Leu Glu Leu Gly Cys Asp Ala Ser Arg Ser Ala Ala Val Gln Asn
                325                 330                 335

Ile Leu Gly Gly Gln Gly Lys Phe Leu Ser Thr Ala Ile Asp Lys Lys
            340                 345                 350

Gly His Ser Ala Trp Ser Gln Val Leu Arg Ala Thr Ala Ser Leu Ala
        355                 360                 365
```

```
Ala His Arg Val Pro Gly Ile Ser Ile Leu Asp Leu Phe His Pro Asn
    370                 375                 380

Phe Arg Glu Met Cys Cys Thr Met Ala Thr Thr Pro Lys Val Glu Asp
385                 390                 395                 400

Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu Lys Glu Met
                405                 410                 415

Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro Ala Pro Ser Glu
            420                 425                 430

Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile Arg Ser Ala Ala Ala
        435                 440                 445

Arg Ser Gly Gln Ser His Asp Cys Ala Ser His Ser His Glu Glu Asn
    450                 455                 460

Lys Asp Ser Cys Pro Glu Lys Leu Lys Leu Asp Ser Val Ser Val Ala
465                 470                 475                 480

Ile Asn Phe Asp Asn Asp Asp Arg Ile Gln Leu Gly His Ala Gly Phe
                485                 490                 495

Arg Glu Met Tyr
            500
```

<210> 59
<211> 1305
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1305)

<400> 59

```
aat aca aga tat agc ttg tac aca ggg gcg atg gca aag gga att gca       48
Asn Thr Arg Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
1               5                   10                  15

tct gca gat ctt gtc att gcc gct ggg aaa gag ggc atc cta gct tcc       96
Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser
            20                  25                  30

tat gga gct gga gga cta cct ctt gct act gtt cga aag gga ata gac      144
Tyr Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
        35                  40                  45

aaa att caa caa gcc ttg cca agt ggc cca tat gct gta aat ctt att      192
Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu Ile
    50                  55                  60

cac tct ccc ttt gac ggc aac ttg gag cag gga aac gtc gat ttg ttc      240
His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp Leu Phe
65                  70                  75                  80

ttg gaa aag aac gtc cgc gtg gcg gaa tgt tcc gcg ttt aca acg cta      288
```

```
Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe Thr Thr Leu
                85                  90                  95

aca gtg cca gta gta cac tat cgt gct gca ggg ctt gtt cgg cgc caa      336
Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu Val Arg Arg Gln
            100                 105                 110

gat gga agc att ttg atc aag aac cga atc att gct aaa gta tct agg      384
Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile Ala Lys Val Ser Arg
        115                 120                 125

aca gaa ctc gct gag atg ttc ctt cgt ccg gca cct caa atc atc ctc      432
Thr Glu Leu Ala Glu Met Phe Leu Arg Pro Ala Pro Gln Ile Ile Leu
    130                 135                 140

gaa aaa ctg gta gca gca gaa atc att tca tct gac caa gcg cgt atg      480
Glu Lys Leu Val Ala Ala Glu Ile Ile Ser Ser Asp Gln Ala Arg Met
145                 150                 155                 160

gca gcc aaa gtt ccc atg gcg gac gac atc gca gtc gaa gcc gac tct      528
Ala Ala Lys Val Pro Met Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
                165                 170                 175

ggt ggg cac acg gat aat cgg cct atg cac gtc att ttg ccc ctg ata      576
Gly Gly His Thr Asp Asn Arg Pro Met His Val Ile Leu Pro Leu Ile
            180                 185                 190

att caa ctc cgc aat act ata ctt gca gag tat ggc tgt gcc acg gct      624
Ile Gln Leu Arg Asn Thr Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala
        195                 200                 205

ttt cgt acc cgt ata ggc gct gga gga ggc att ggt tgt cct tca gcg      672
Phe Arg Thr Arg Ile Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala
    210                 215                 220

gcc ctc gca gcc ttt gat atg ggt gcg agt ttt gtc gtg act gga agc      720
Ala Leu Ala Ala Phe Asp Met Gly Ala Ser Phe Val Val Thr Gly Ser
225                 230                 235                 240

ata aat caa att tgc cgc gag gca ggg act tgc gat act gtt cgg gag      768
Ile Asn Gln Ile Cys Arg Glu Ala Gly Thr Cys Asp Thr Val Arg Glu
                245                 250                 255

cta ctt gcc aac tca agc tac tcg gac gtg acg atg gcg cca gca gca      816
Leu Leu Ala Asn Ser Ser Tyr Ser Asp Val Thr Met Ala Pro Ala Ala
            260                 265                 270

gac atg ttt gac caa ggt gtg aaa ctc caa gtc tta aaa cga gga acg      864
Asp Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr
        275                 280                 285

atg ttt cca agc aga gca aat aaa ctc cgg aag ctc ttt gtg aac tac      912
Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn Tyr
    290                 295                 300

gaa tct cta gaa aca ctc ccg tcg aaa gag ttg aaa tac ctg gaa aac      960
Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu Glu Asn
305                 310                 315                 320

atc ata ttc aag caa gca gta gac cag gtg tgg gag gaa aca aag cgc     1008
Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu Thr Lys Arg
                325                 330                 335

ttt tac tgt gaa aaa ctg aac aat cca gat aaa att gca agg gcc atg     1056
Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile Ala Arg Ala Met
            340                 345                 350

aaa gat cct aaa ttg aag atg tcg ctt tgc ttt cgg tgg tat ctc tcc     1104
```

```
Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
        355                 360                 365

aag agc tct ggg tgg gcc aac gca gga att aaa tct cgt gca ctc gac     1152
Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile Lys Ser Arg Ala Leu Asp
    370                 375                 380

tac cag atc tgg tgt ggc ccg gca atg ggc tcg ttc aac aat ttc gcc     1200
Tyr Gln Ile Trp Cys Gly Pro Ala Met Gly Ser Phe Asn Asn Phe Ala
385                 390                 395                 400

agc ggc aca tcc ctc gat tgg aaa gtg act ggg gtt ttc cct ggc gtt     1248
Ser Gly Thr Ser Leu Asp Trp Lys Val Thr Gly Val Phe Pro Gly Val
                405                 410                 415

gcg gaa gta aac atg gcc att tta gat ggc gcg cga gaa cta gct gct     1296
Ala Glu Val Asn Met Ala Ile Leu Asp Gly Ala Arg Glu Leu Ala Ala
            420                 425                 430

aaa cga aat                                                        1305
Lys Arg Asn
        435


<210>  60
<211>  435
<212>  PRT
<213>  Thraustochytrium sp.

<400>  60

Asn Thr Arg Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
1               5                   10                  15


Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser
            20                  25                  30


Tyr Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
        35                  40                  45


Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu Ile
    50                  55                  60


His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp Leu Phe
65                  70                  75                  80


Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe Thr Thr Leu
                85                  90                  95


Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu Val Arg Arg Gln
            100                 105                 110


Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile Ala Lys Val Ser Arg
        115                 120                 125


Thr Glu Leu Ala Glu Met Phe Leu Arg Pro Ala Pro Gln Ile Ile Leu
    130                 135                 140


Glu Lys Leu Val Ala Ala Glu Ile Ile Ser Ser Asp Gln Ala Arg Met
```

145 150 155 160

Ala Ala Lys Val Pro Met Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
165 170 175

Gly Gly His Thr Asp Asn Arg Pro Met His Val Ile Leu Pro Leu Ile
180 185 190

Ile Gln Leu Arg Asn Thr Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala
195 200 205

Phe Arg Thr Arg Ile Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala
210 215 220

Ala Leu Ala Ala Phe Asp Met Gly Ala Ser Phe Val Val Thr Gly Ser
225 230 235 240

Ile Asn Gln Ile Cys Arg Glu Ala Gly Thr Cys Asp Thr Val Arg Glu
245 250 255

Leu Leu Ala Asn Ser Ser Tyr Ser Asp Val Thr Met Ala Pro Ala Ala
260 265 270

Asp Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr
275 280 285

Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn Tyr
290 295 300

Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu Glu Asn
305 310 315 320

Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu Thr Lys Arg
325 330 335

Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile Ala Arg Ala Met
340 345 350

Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
355 360 365

Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile Lys Ser Arg Ala Leu Asp
370 375 380

Tyr Gln Ile Trp Cys Gly Pro Ala Met Gly Ser Phe Asn Asn Phe Ala
385 390 395 400

Ser Gly Thr Ser Leu Asp Trp Lys Val Thr Gly Val Phe Pro Gly Val
405 410 415

Ala Glu Val Asn Met Ala Ile Leu Asp Gly Ala Arg Glu Leu Ala Ala

```
            420                 425                 430

Lys Arg Asn
        435


<210>  61
<211>  4410
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(4410)

<400>  61
atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc       48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1               5                  10                  15

aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag       96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
            20                  25                  30

ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag      144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
        35                  40                  45

ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga      192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

gag tac ctg ctg gtg acc cgg gtt acg ctg atg gat gcc gag gtg ggc      240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65                  70                  75                  80

aac ttt cga gtg gga gca cgt atg gtg aca gag tat gac gta cct gtg      288
Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
                85                  90                  95

aac gga gag ctc tcg gaa ggg gga gat gtg ccg tgg gct gtg ttg gtg      336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
            100                 105                 110

gaa gcc ggg cag tgc gac ttg ctg cta att tct tac atg ggc atc gat      384
Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

ttc cag tgc aaa gga gag cgg gtc tac cgg ctg ctg aac acc acc ttg      432
Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

acg ttt ttt ggc gtc gcg aaa gaa ggg gaa acg ctt gtg tac gat att      480
Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145                 150                 155                 160

cgc gtc acg ggt ttc gcc aag agg ccg gac gga gat atc tcc atg ttc      528
Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
                165                 170                 175

ttt ttc gaa tat gat tgc tac tgc aat ggc aag ctt ctc atc gaa atg      576
Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
            180                 185                 190

cga gat ggc tct gca ggc ttc ttc acg gac gaa gag ctc gct gcc ggc      624
Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
```

```
        195                 200                 205

aaa gga gtg gtc gtc act cgt gca cag caa aac atg cgg gac aaa att      672
Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
    210                 215                 220

gta cgg cag tcc att gag cct ttt gca ctg gcg gct tgc acg cac aaa      720
Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225                 230                 235                 240

acg act ctg aac gag agt gac atg cag tcc ctt gtg gag cga aac tgg      768
Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
                245                 250                 255

gca aac gtt ttt ggc acc agt aac aag atg gcg gag ctc aac tat aaa      816
Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
            260                 265                 270

att tgc gcc agg aaa atg ctc atg atc gac agg gtt acc cac att gac      864
Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
        275                 280                 285

cac cac ggt ggg gcg tat ggc ctc gga cta ctt gtt gga gag aag atc      912
His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
    290                 295                 300

ttg gat cga aac cat tgg tac ttt cct tgt cac ttt gtc aat gat caa      960
Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305                 310                 315                 320

gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa     1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
                325                 330                 335

ctc tat atg atc tgg ctt ggc ctc cac ctg aaa atg gag gaa ttt gat     1056
Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
            340                 345                 350

ttt ctc cca gtt agc ggc cac aaa aac aag gtg cga tgc agg gga caa     1104
Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

att tca ccg cat aaa ggc aag ctt gtc tac gtc atg gaa atc aaa aag     1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
    370                 375                 380

atg ggt tac gat caa gca tct gga agc cca tac gcc atc gcg gac gtt     1200
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385                 390                 395                 400

gat atc att gac gtc aac gaa gag ctg ggt caa agt ttt gac atc aac     1248
Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
                405                 410                 415

gac ctt gcg agc tac gga aaa ggt gac ctg agc aaa aaa atc gtg gtt     1296
Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
            420                 425                 430

gac ttc aaa gga att gct ttg cag ctc aaa ggc cgc gct ttt tca cgc     1344
Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
        435                 440                 445

atg agt tcc agc tcg tcc ttg aac gaa gga tgg caa tgt gtt cca aaa     1392
Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
    450                 455                 460

cca agc cag aga atg gaa cac gaa cag ccc cct gct cac tgc ctt gca     1440
Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
```

```
465                 470                 475                 480

agc gac ccc gaa gcc cct tca act gtg acc tgg cac cca atg tca aag     1488
Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
                485                 490                 495

ctt cct ggc aac cct acg ccg ttc ttc tcc cct tca tct tac cct ccg     1536
Leu Pro Gly Asn Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro
            500                 505                 510

agg gca att tgc ttc atc cct ttc ccg ggc aat ccc ctt gac aac aac     1584
Arg Ala Ile Cys Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn
        515                 520                 525

tgc aag gct gga gaa atg ccc ctg aac tgg tac aac atg tca gag ttc     1632
Cys Lys Ala Gly Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe
    530                 535                 540

atg tgt ggc aag gtt tct aac tgc ttg ggc cca gaa ttc gca cgc ttt     1680
Met Cys Gly Lys Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe
545                 550                 555                 560

gac aag tcg aac acc agc cgg agc cct gct ttt gac ttg gct ctg gtg     1728
Asp Lys Ser Asn Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val
                565                 570                 575

acc cga gtt gtt gaa gtc aca aac atg gaa cac ggc aag ttt cta aac     1776
Thr Arg Val Val Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn
            580                 585                 590

gtt gat tgc aat cca agc aaa ggc aca atg gtg ggg gag ttt gac tgt     1824
Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys
        595                 600                 605

ccc caa gac gcg tgg ttc ttt gat ggt tcg tgc aac gac ggc cat atg     1872
Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met
    610                 615                 620

ccg tat tcc att atc atg gaa atc gga ctg caa acc tca ggt gtt ctc     1920
Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu
625                 630                 635                 640

acc tcg gtg ttg aag gca ccg ctg act atg gac aag gat gac att ctc     1968
Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu
                645                 650                 655

ttt cga aac ctc gat gca agt gct gaa atg gtg cgt cca gac gtg gat     2016
Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp
            660                 665                 670

gtt cgc ggc aaa acg att cga aac gtg acc aag tgt acc ggc tat gca     2064
Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala
        675                 680                 685

atg ttg gga aag atg ggg att cac cgg ttc acg ttt gag ttg agc gtt     2112
Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val
    690                 695                 700

gac ggc gtg gta ttt tat aaa gga tcc act tcc ttt gga tgg ttc act     2160
Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr
705                 710                 715                 720

ccc gag gtg ttt gct cag caa gct gga ctc gac aac ggg aaa aag acg     2208
Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr
                725                 730                 735

gag ccc tgg tgc aag act aac aac acc tcg gtt cga aga gtt gaa atc     2256
Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile
```

```
            740                 745                 750

gca tcc gcc aaa gga aaa gag cag ctg act gag aag ctt ccc gac gca     2304
Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala
        755                 760                 765

act aat gct caa gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat     2352
Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp
    770                 775                 780

tac ctc aat att gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc     2400
Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala
785                 790                 795                 800

cac gga cac aaa gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac     2448
His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His
                805                 810                 815

ttt tgg ttc gat cct gta atg cca gga tct tta gga att gaa tca atg     2496
Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met
            820                 825                 830

ttc cag ctt atc gag gcc ttt gcg gtg gac caa aac att cct gga gag     2544
Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu
        835                 840                 845

tac aac gta tcc aat ccg acc ttt gcc cat gca cca ggc aaa acg gcg     2592
Tyr Asn Val Ser Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala
    850                 855                 860

tgg aaa tac cga ggc cag ctc aca cca aag aac cgt gcg atg gac tgc     2640
Trp Lys Tyr Arg Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys
865                 870                 875                 880

gag gtg cat atc gtt tca att acc gcc tcc ccc gag aac ggg ggc tac     2688
Glu Val His Ile Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr
                885                 890                 895

gtt gac atc gtg gcc gat gga gcg ctt tgg gta gat gga ctt cgc gtg     2736
Val Asp Ile Val Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val
            900                 905                 910

tac gaa gcc aaa gag ctt cga gtt cgt gtc gtt tcg gca aaa cct caa     2784
Tyr Glu Ala Lys Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln
        915                 920                 925

gca att ccg gat gta caa caa cag cca cct agc gca aag gcg gac ccg     2832
Ala Ile Pro Asp Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro
    930                 935                 940

ggg aaa aca gga gtt gca ctt tcg ccc act cag cta cgc gac gtc ctg     2880
Gly Lys Thr Gly Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu
945                 950                 955                 960

ctt gaa gtg gac aat cca ttg tat ctt ggt gta gag aac tcc aat ttg     2928
Leu Glu Val Asp Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu
                965                 970                 975

gtg cag ttt gag tcg aaa cct gca act tct tca cgt atc gtt tcg atc     2976
Val Gln Phe Glu Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile
            980                 985                 990

aaa ccg tgc tcg att agt gac ctt  ggc gat aag tct ttt  atg gaa acg   3024
Lys Pro Cys Ser Ile Ser Asp Leu  Gly Asp Lys Ser Phe  Met Glu Thr
        995                 1000                1005

tac aac  gtg tca gca cct ctg  tat act gga gca atg  gcc aag ggc      3069
Tyr Asn  Val Ser Ala Pro Leu  Tyr Thr Gly Ala Met  Ala Lys Gly
```

```
    1010                1015                1020

att gca  tcc gcc gac ttg gtc  att gct gct ggg aaa  cgc aag ata      3114
Ile Ala  Ser Ala Asp Leu Val  Ile Ala Ala Gly Lys  Arg Lys Ile
    1025                1030                1035

ctt gga  tcg ttt ggt gcg gga  ggg ctg cct att tcc  ata gtc cgt      3159
Leu Gly  Ser Phe Gly Ala Gly  Gly Leu Pro Ile Ser  Ile Val Arg
    1040                1045                1050

gaa gca  ctg gag aaa att caa  caa cac ctg ccc cac  ggc ccc tac      3204
Glu Ala  Leu Glu Lys Ile Gln  Gln His Leu Pro His  Gly Pro Tyr
    1055                1060                1065

gct gtt  aac ctc att cac tcg  cct ttc gac agc aac  ttg gaa aag      3249
Ala Val  Asn Leu Ile His Ser  Pro Phe Asp Ser Asn  Leu Glu Lys
    1070                1075                1080

ggc aac  gtt gac ctc ttt ctc  gag atg ggc gtg aca  gtg gta gaa      3294
Gly Asn  Val Asp Leu Phe Leu  Glu Met Gly Val Thr  Val Val Glu
    1085                1090                1095

tgc agc  gcg ttc atg gaa ctc  acg gcc cag gtt gtc  cgg tac cgc      3339
Cys Ser  Ala Phe Met Glu Leu  Thr Ala Gln Val Val  Arg Tyr Arg
    1100                1105                1110

gcg tct  ggt cta agc aaa agt  gcg gac ggt tcg att  cgc att gct      3384
Ala Ser  Gly Leu Ser Lys Ser  Ala Asp Gly Ser Ile  Arg Ile Ala
    1115                1120                1125

cac cgt  att att ggc aag gtt  tcc aga acc gag ctg  gca gaa atg      3429
His Arg  Ile Ile Gly Lys Val  Ser Arg Thr Glu Leu  Ala Glu Met
    1130                1135                1140

ttt att  cgt cca gca cca cag  cac ctc ctc caa aaa  ctc gta gcc      3474
Phe Ile  Arg Pro Ala Pro Gln  His Leu Leu Gln Lys  Leu Val Ala
    1145                1150                1155

tcc ggc  gag ctg aca gct gag  caa gcc gag ctt gca  aca cag gtt      3519
Ser Gly  Glu Leu Thr Ala Glu  Gln Ala Glu Leu Ala  Thr Gln Val
    1160                1165                1170

ccg gtg  gcg gat gac att gcg  gtc gaa gcc gac tcg  ggg ggg cat      3564
Pro Val  Ala Asp Asp Ile Ala  Val Glu Ala Asp Ser  Gly Gly His
    1175                1180                1185

acc gac  aac agg cct att cac  gtc att ctt cct cta  atc atc aac      3609
Thr Asp  Asn Arg Pro Ile His  Val Ile Leu Pro Leu  Ile Ile Asn
    1190                1195                1200

cta cgc  aac cgt ttg cat aaa  gag ctt gac tac cct  tcg cat ctc      3654
Leu Arg  Asn Arg Leu His Lys  Glu Leu Asp Tyr Pro  Ser His Leu
    1205                1210                1215

cgg gta  cgt gtg ggt gct ggt  ggt ggt att gga tgt  cct caa gcc      3699
Arg Val  Arg Val Gly Ala Gly  Gly Gly Ile Gly Cys  Pro Gln Ala
    1220                1225                1230

gct ctt  gca gca ttt caa atg  ggg gca gcg ttt tta  atc act gga      3744
Ala Leu  Ala Ala Phe Gln Met  Gly Ala Ala Phe Leu  Ile Thr Gly
    1235                1240                1245

acg gtg  aac cag ctt gct cgt  gaa agt ggc act tgt  gac aac gtc      3789
Thr Val  Asn Gln Leu Ala Arg  Glu Ser Gly Thr Cys  Asp Asn Val
    1250                1255                1260

cgg tta  cag ctc tca aag gcc  acg tat agc gac gtg  tgt atg gct      3834
Arg Leu  Gln Leu Ser Lys Ala  Thr Tyr Ser Asp Val  Cys Met Ala
```

```
    1265                1270                1275

cct gct  gcc gat atg ttt gac  caa ggc gtg gag ctg  caa gta ttg      3879
Pro Ala  Ala Asp Met Phe Asp  Gln Gly Val Glu Leu  Gln Val Leu
    1280                1285                1290

aag aaa  ggc acg ctg ttc cca  agt cgt gct aag aag  ctg tac gag      3924
Lys Lys  Gly Thr Leu Phe Pro  Ser Arg Ala Lys Lys  Leu Tyr Glu
    1295                1300                1305

ctg ttc  tgc aag tat gac tcg  ttt gag gca atg ccg  gct gaa gaa      3969
Leu Phe  Cys Lys Tyr Asp Ser  Phe Glu Ala Met Pro  Ala Glu Glu
    1310                1315                1320

ttg caa  cgg gtt gaa aag cgg  att ttt caa aag tcg  ctt gct gaa      4014
Leu Gln  Arg Val Glu Lys Arg  Ile Phe Gln Lys Ser  Leu Ala Glu
    1325                1330                1335

gtt tgg  cag gag acc agt gac  ttt tac att cat cgt  atc aag aac      4059
Val Trp  Gln Glu Thr Ser Asp  Phe Tyr Ile His Arg  Ile Lys Asn
    1340                1345                1350

cct gag  aaa atc aat cgt gct  gca agc gat ggc aaa  ctg aaa atg      4104
Pro Glu  Lys Ile Asn Arg Ala  Ala Ser Asp Gly Lys  Leu Lys Met
    1355                1360                1365

tcg ctt  tgc ttt cgc tgg tac  ctt ggg ctt tcc tca  ttt tgg gcc      4149
Ser Leu  Cys Phe Arg Trp Tyr  Leu Gly Leu Ser Ser  Phe Trp Ala
    1370                1375                1380

aac tct  ggg gca caa gat cgc  gtc atg gac tat caa  att tgg tgt      4194
Asn Ser  Gly Ala Gln Asp Arg  Val Met Asp Tyr Gln  Ile Trp Cys
    1385                1390                1395

ggc cct  gct att ggc gct ttc  aat gat ttt acc aag  ggc acg tac      4239
Gly Pro  Ala Ile Gly Ala Phe  Asn Asp Phe Thr Lys  Gly Thr Tyr
    1400                1405                1410

ctt gac  gtg act gtt gca aag  agt tac cct tgt gtg  gca cag atc      4284
Leu Asp  Val Thr Val Ala Lys  Ser Tyr Pro Cys Val  Ala Gln Ile
    1415                1420                1425

aat ttg  caa att ttg caa gga  gct gcg tat ctg aaa  cgc ctt ggt      4329
Asn Leu  Gln Ile Leu Gln Gly  Ala Ala Tyr Leu Lys  Arg Leu Gly
    1430                1435                1440

gtc att  cgt ttt gac cgc atg  ctg ctg cag gcc gtc  gat atc gac      4374
Val Ile  Arg Phe Asp Arg Met  Leu Leu Gln Ala Val  Asp Ile Asp
    1445                1450                1455

gat cct  gta ttt act tac gtg  ccg acc cag cca ctt                  4410
Asp Pro  Val Phe Thr Tyr Val  Pro Thr Gln Pro Leu
    1460                1465                1470


<210>  62
<211>  1470
<212>  PRT
<213>  Thraustochytrium sp.

<400>  62

Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1               5                  10                  15


Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
            20                  25                  30
```

```
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
        35                  40                  45

Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65                  70                  75                  80

Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
                85                  90                  95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
            100                 105                 110

Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145                 150                 155                 160

Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
                165                 170                 175

Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
            180                 185                 190

Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
        195                 200                 205

Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
    210                 215                 220

Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225                 230                 235                 240

Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
                245                 250                 255

Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
            260                 265                 270

Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
        275                 280                 285

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
    290                 295                 300
```

```
Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305                 310                 315                 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
                325                 330                 335

Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
            340                 345                 350

Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
    370                 375                 380

Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385                 390                 395                 400

Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
                405                 410                 415

Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
            420                 425                 430

Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
        435                 440                 445

Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
    450                 455                 460

Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465                 470                 475                 480

Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
                485                 490                 495

Leu Pro Gly Asn Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro
            500                 505                 510

Arg Ala Ile Cys Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn
        515                 520                 525

Cys Lys Ala Gly Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe
    530                 535                 540

Met Cys Gly Lys Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe
545                 550                 555                 560

Asp Lys Ser Asn Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val
                565                 570                 575
```

```
Thr Arg Val Val Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn
            580                 585                 590

Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys
        595                 600                 605

Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met
    610                 615                 620

Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu
625                 630                 635                 640

Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu
                645                 650                 655

Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp
            660                 665                 670

Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala
        675                 680                 685

Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val
    690                 695                 700

Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr
705                 710                 715                 720

Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr
                725                 730                 735

Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile
            740                 745                 750

Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala
        755                 760                 765

Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp
    770                 775                 780

Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala
785                 790                 795                 800

His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His
                805                 810                 815

Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met
            820                 825                 830

Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu
        835                 840                 845
```

```
Tyr Asn Val Ser Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala
    850                 855                 860

Trp Lys Tyr Arg Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys
865                 870                 875                 880

Glu Val His Ile Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr
                885                 890                 895

Val Asp Ile Val Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val
            900                 905                 910

Tyr Glu Ala Lys Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln
        915                 920                 925

Ala Ile Pro Asp Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro
    930                 935                 940

Gly Lys Thr Gly Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu
945                 950                 955                 960

Leu Glu Val Asp Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu
                965                 970                 975

Val Gln Phe Glu Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile
            980                 985                 990

Lys Pro Cys Ser Ile Ser Asp Leu  Gly Asp Lys Ser Phe  Met Glu Thr
        995                 1000                1005

Tyr Asn  Val Ser Ala Pro Leu  Tyr Thr Gly Ala Met  Ala Lys Gly
    1010                1015                1020

Ile Ala  Ser Ala Asp Leu Val  Ile Ala Ala Gly Lys  Arg Lys Ile
    1025                1030                1035

Leu Gly  Ser Phe Gly Ala Gly  Gly Leu Pro Ile Ser  Ile Val Arg
    1040                1045                1050

Glu Ala  Leu Glu Lys Ile Gln  Gln His Leu Pro His  Gly Pro Tyr
    1055                1060                1065

Ala Val  Asn Leu Ile His Ser  Pro Phe Asp Ser Asn  Leu Glu Lys
    1070                1075                1080

Gly Asn  Val Asp Leu Phe Leu  Glu Met Gly Val Thr  Val Val Glu
    1085                1090                1095

Cys Ser  Ala Phe Met Glu Leu  Thr Ala Gln Val Val  Arg Tyr Arg
    1100                1105                1110
```

```
Ala Ser  Gly Leu Ser Lys Ser  Ala Asp Gly Ser Ile  Arg Ile Ala
    1115                 1120                 1125

His Arg  Ile Ile Gly Lys Val  Ser Arg Thr Glu Leu  Ala Glu Met
    1130                 1135                 1140

Phe Ile  Arg Pro Ala Pro Gln  His Leu Leu Gln Lys  Leu Val Ala
    1145                 1150                 1155

Ser Gly  Glu Leu Thr Ala Glu  Gln Ala Glu Leu Ala  Thr Gln Val
    1160                 1165                 1170

Pro Val  Ala Asp Asp Ile Ala  Val Glu Ala Asp Ser  Gly Gly His
    1175                 1180                 1185

Thr Asp  Asn Arg Pro Ile His  Val Ile Leu Pro Leu  Ile Ile Asn
    1190                 1195                 1200

Leu Arg  Asn Arg Leu His Lys  Glu Leu Asp Tyr Pro  Ser His Leu
    1205                 1210                 1215

Arg Val  Arg Val Gly Ala Gly  Gly Gly Ile Gly Cys  Pro Gln Ala
    1220                 1225                 1230

Ala Leu  Ala Ala Phe Gln Met  Gly Ala Ala Phe Leu  Ile Thr Gly
    1235                 1240                 1245

Thr Val  Asn Gln Leu Ala Arg  Glu Ser Gly Thr Cys  Asp Asn Val
    1250                 1255                 1260

Arg Leu  Gln Leu Ser Lys Ala  Thr Tyr Ser Asp Val  Cys Met Ala
    1265                 1270                 1275

Pro Ala  Ala Asp Met Phe Asp  Gln Gly Val Glu Leu  Gln Val Leu
    1280                 1285                 1290

Lys Lys  Gly Thr Leu Phe Pro  Ser Arg Ala Lys Lys  Leu Tyr Glu
    1295                 1300                 1305

Leu Phe  Cys Lys Tyr Asp Ser  Phe Glu Ala Met Pro  Ala Glu Glu
    1310                 1315                 1320

Leu Gln  Arg Val Glu Lys Arg  Ile Phe Gln Lys Ser  Leu Ala Glu
    1325                 1330                 1335

Val Trp  Gln Glu Thr Ser Asp  Phe Tyr Ile His Arg  Ile Lys Asn
    1340                 1345                 1350

Pro Glu  Lys Ile Asn Arg Ala  Ala Ser Asp Gly Lys  Leu Lys Met
    1355                 1360                 1365
```

```
Ser Leu  Cys Phe Arg Trp Tyr  Leu Gly Leu Ser Ser  Phe Trp Ala
    1370                1375                1380

Asn Ser  Gly Ala Gln Asp Arg  Val Met Asp Tyr Gln  Ile Trp Cys
    1385                1390                1395

Gly Pro  Ala Ile Gly Ala Phe  Asn Asp Phe Thr Lys  Gly Thr Tyr
    1400                1405                1410

Leu Asp  Val Thr Val Ala Lys  Ser Tyr Pro Cys Val  Ala Gln Ile
    1415                1420                1425

Asn Leu  Gln Ile Leu Gln Gly  Ala Ala Tyr Leu Lys  Arg Leu Gly
    1430                1435                1440

Val Ile  Arg Phe Asp Arg Met  Leu Leu Gln Ala Val  Asp Ile Asp
    1445                1450                1455

Asp Pro  Val Phe Thr Tyr Val  Pro Thr Gln Pro Leu
    1460                1465                1470
```

<210> 63
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 63

```
atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc       48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1               5                   10                  15

aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag       96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
            20                  25                  30

ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag      144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
        35                  40                  45

ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga      192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

gag tac ctg ctg gtg acc cgg gtt acg ctg atg gat gcc gag gtg ggc      240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65                  70                  75                  80

aac ttt cga gtg gga gca cgt atg gtg aca gag tat gac gta cct gtg      288
Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
                85                  90                  95

aac gga gag ctc tcg gaa ggg gga gat gtg ccg tgg gct gtg ttg gtg      336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
            100                 105                 110
```

```
gaa gcc ggg cag tgc gac ttg ctg cta att tct tac atg ggc atc gat      384
Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

ttc cag tgc aaa gga gag cgg gtc tac cgg ctg ctg aac acc acc ttg      432
Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

acg ttt ttt ggc gtc gcg aaa gaa ggg gaa acg ctt gtg tac gat att      480
Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145                 150                 155                 160

cgc gtc acg ggt ttc gcc aag agg ccg gac gga gat atc tcc atg ttc      528
Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
                165                 170                 175

ttt ttc gaa tat gat tgc tac tgc aat ggc aag ctt ctc atc gaa atg      576
Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
            180                 185                 190

cga gat ggc tct gca ggc ttc ttc acg gac gaa gag ctc gct gcc ggc      624
Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
        195                 200                 205

aaa gga gtg gtc gtc act cgt gca cag caa aac atg cgg gac aaa att      672
Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
    210                 215                 220

gta cgg cag tcc att gag cct ttt gca ctg gcg gct tgc acg cac aaa      720
Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225                 230                 235                 240

acg act ctg aac gag agt gac atg cag tcc ctt gtg gag cga aac tgg      768
Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
                245                 250                 255

gca aac gtt ttt ggc acc agt aac aag atg gcg gag ctc aac tat aaa      816
Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
            260                 265                 270

att tgc gcc agg aaa atg ctc atg atc gac agg gtt acc cac att gac      864
Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
        275                 280                 285

cac cac ggt ggg gcg tat ggc ctc gga cta ctt gtt gga gag aag atc      912
His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
    290                 295                 300

ttg gat cga aac cat tgg tac ttt cct tgt cac ttt gtc aat gat caa      960
Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305                 310                 315                 320

gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa     1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
                325                 330                 335

ctc tat atg atc tgg ctt ggc ctc cac ctg aaa atg gag gaa ttt gat     1056
Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
            340                 345                 350

ttt ctc cca gtt agc ggc cac aaa aac aag gtg cga tgc agg gga caa     1104
Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

att tca ccg cat aaa ggc aag ctt gtc tac gtc atg gaa atc aaa aag     1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
    370                 375                 380
```

```
atg ggt tac gat caa gca tct gga agc cca tac gcc atc gcg gac gtt     1200
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385                 390                 395                 400

gat atc att gac gtc aac gaa gag ctg ggt caa agt ttt gac atc aac     1248
Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
                405                 410                 415

gac ctt gcg agc tac gga aaa ggt gac ctg agc aaa aaa atc gtg gtt     1296
Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
            420                 425                 430

gac ttc aaa gga att gct ttg cag ctc aaa ggc cgc gct ttt tca cgc     1344
Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
        435                 440                 445

atg agt tcc agc tcg tcc ttg aac gaa gga tgg caa tgt gtt cca aaa     1392
Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
    450                 455                 460

cca agc cag aga atg gaa cac gaa cag ccc cct gct cac tgc ctt gca     1440
Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465                 470                 475                 480

agc gac ccc gaa gcc cct tca act gtg acc tgg cac cca atg tca aag     1488
Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
                485                 490                 495

ctt cct ggc aac                                                    1500
Leu Pro Gly Asn
            500
```

<210> 64
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 64

```
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1               5                   10                  15

Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
            20                  25                  30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
        35                  40                  45

Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65                  70                  75                  80

Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
                85                  90                  95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
            100                 105                 110
```

```
Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145                 150                 155                 160

Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
                165                 170                 175

Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
            180                 185                 190

Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
        195                 200                 205

Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
    210                 215                 220

Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225                 230                 235                 240

Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
                245                 250                 255

Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
            260                 265                 270

Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
        275                 280                 285

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
    290                 295                 300

Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305                 310                 315                 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
                325                 330                 335

Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
            340                 345                 350

Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
    370                 375                 380
```

```
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385                 390                 395                 400


Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
                405                 410                 415


Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
            420                 425                 430


Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
        435                 440                 445


Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
    450                 455                 460


Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465                 470                 475                 480


Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
                485                 490                 495


Leu Pro Gly Asn
            500


<210>  65
<211>  1500
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1500)

<400>  65
cct acg ccg ttc ttc tcc cct tca tct tac cct ccg agg gca att tgc       48
Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys
1               5                   10                  15

ttc atc cct ttc ccg ggc aat ccc ctt gac aac aac tgc aag gct gga       96
Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly
            20                  25                  30

gaa atg ccc ctg aac tgg tac aac atg tca gag ttc atg tgt ggc aag      144
Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys
        35                  40                  45

gtt tct aac tgc ttg ggc cca gaa ttc gca cgc ttt gac aag tcg aac      192
Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn
    50                  55                  60

acc agc cgg agc cct gct ttt gac ttg gct ctg gtg acc cga gtt gtt      240
Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val
65                  70                  75                  80

gaa gtc aca aac atg gaa cac ggc aag ttt cta aac gtt gat tgc aat      288
Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn
                85                  90                  95
```

```
cca agc aaa ggc aca atg gtg ggg gag ttt gac tgt ccc caa gac gcg      336
Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala
            100                 105                 110

tgg ttc ttt gat ggt tcg tgc aac gac ggc cat atg ccg tat tcc att      384
Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile
        115                 120                 125

atc atg gaa atc gga ctg caa acc tca ggt gtt ctc acc tcg gtg ttg      432
Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu
    130                 135                 140

aag gca ccg ctg act atg gac aag gat gac att ctc ttt cga aac ctc      480
Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu
145                 150                 155                 160

gat gca agt gct gaa atg gtg cgt cca gac gtg gat gtt cgc ggc aaa      528
Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys
                165                 170                 175

acg att cga aac gtg acc aag tgt acc ggc tat gca atg ttg gga aag      576
Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys
            180                 185                 190

atg ggg att cac cgg ttc acg ttt gag ttg agc gtt gac ggc gtg gta      624
Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val
        195                 200                 205

ttt tat aaa gga tcc act tcc ttt gga tgg ttc act ccc gag gtg ttt      672
Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe
    210                 215                 220

gct cag caa gct gga ctc gac aac ggg aaa aag acg gag ccc tgg tgc      720
Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys
225                 230                 235                 240

aag act aac aac acc tcg gtt cga aga gtt gaa atc gca tcc gcc aaa      768
Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys
                245                 250                 255

gga aaa gag cag ctg act gag aag ctt ccc gac gca act aat gct caa      816
Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln
            260                 265                 270

gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat tac ctc aat att      864
Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile
        275                 280                 285

gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc cac gga cac aaa      912
Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys
    290                 295                 300

gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac ttt tgg ttc gat      960
Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp
305                 310                 315                 320

cct gta atg cca gga tct tta gga att gaa tca atg ttc cag ctt atc     1008
Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile
                325                 330                 335

gag gcc ttt gcg gtg gac caa aac att cct gga gag tac aac gta tcc     1056
Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser
            340                 345                 350

aat ccg acc ttt gcc cat gca cca ggc aaa acg gcg tgg aaa tac cga     1104
Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg
        355                 360                 365
```

```
ggc cag ctc aca cca aag aac cgt gcg atg gac tgc gag gtg cat atc     1152
Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile
    370                 375                 380

gtt tca att acc gcc tcc ccc gag aac ggg ggc tac gtt gac atc gtg     1200
Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val
385                 390                 395                 400

gcc gat gga gcg ctt tgg gta gat gga ctt cgc gtg tac gaa gcc aaa     1248
Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys
                405                 410                 415

gag ctt cga gtt cgt gtc gtt tcg gca aaa cct caa gca att ccg gat     1296
Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp
            420                 425                 430

gta caa caa cag cca cct agc gca aag gcg gac ccg ggg aaa aca gga     1344
Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly
        435                 440                 445

gtt gca ctt tcg ccc act cag cta cgc gac gtc ctg ctt gaa gtg gac     1392
Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp
    450                 455                 460

aat cca ttg tat ctt ggt gta gag aac tcc aat ttg gtg cag ttt gag     1440
Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu
465                 470                 475                 480

tcg aaa cct gca act tct tca cgt atc gtt tcg atc aaa ccg tgc tcg     1488
Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser
                485                 490                 495

att agt gac ctt                                                    1500
Ile Ser Asp Leu
            500
```

<210> 66
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 66

```
Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys
1               5                   10                  15

Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly
            20                  25                  30

Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys
        35                  40                  45

Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn
    50                  55                  60

Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val
65                  70                  75                  80

Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn
                85                  90                  95
```

```
Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala
            100                 105                 110

Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile
        115                 120                 125

Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu
    130                 135                 140

Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu
145                 150                 155                 160

Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys
                165                 170                 175

Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys
            180                 185                 190

Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val
        195                 200                 205

Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe
    210                 215                 220

Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys
225                 230                 235                 240

Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys
                245                 250                 255

Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln
            260                 265                 270

Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile
        275                 280                 285

Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys
    290                 295                 300

Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp
305                 310                 315                 320

Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile
                325                 330                 335

Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser
            340                 345                 350

Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg
        355                 360                 365
```

```
Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile
    370                 375                 380

Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val
385                 390                 395                 400

Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys
                405                 410                 415

Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp
            420                 425                 430

Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly
        435                 440                 445

Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp
    450                 455                 460

Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu
465                 470                 475                 480

Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser
                485                 490                 495

Ile Ser Asp Leu
            500


<210>  67
<211>  1410
<212>  DNA
<213>  Thraustochytrium sp.


<220>
<221>  CDS
<222>  (1)..(1410)

<400>  67
ggc gat aag tct ttt atg gaa acg tac aac gtg tca gca cct ctg tat       48
Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala Pro Leu Tyr
1               5                   10                  15

act gga gca atg gcc aag ggc att gca tcc gcc gac ttg gtc att gct       96
Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala
            20                  25                  30

gct ggg aaa cgc aag ata ctt gga tcg ttt ggt gcg gga ggg ctg cct      144
Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
        35                  40                  45

att tcc ata gtc cgt gaa gca ctg gag aaa att caa caa cac ctg ccc      192
Ile Ser Ile Val Arg Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro
    50                  55                  60

cac ggc ccc tac gct gtt aac ctc att cac tcg cct ttc gac agc aac      240
His Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
65                  70                  75                  80

ttg gaa aag ggc aac gtt gac ctc ttt ctc gag atg ggc gtg aca gtg      288
```

```
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val
                85                  90                  95

gta gaa tgc agc gcg ttc atg gaa ctc acg gcc cag gtt gtc cgg tac      336
Val Glu Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr
            100                 105                 110

cgc gcg tct ggt cta agc aaa agt gcg gac ggt tcg att cgc att gct      384
Arg Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala
        115                 120                 125

cac cgt att att ggc aag gtt tcc aga acc gag ctg gca gaa atg ttt      432
His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe
    130                 135                 140

att cgt cca gca cca cag cac ctc ctc caa aaa ctc gta gcc tcc ggc      480
Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala Ser Gly
145                 150                 155                 160

gag ctg aca gct gag caa gcc gag ctt gca aca cag gtt ccg gtg gcg      528
Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val Pro Val Ala
                165                 170                 175

gat gac att gcg gtc gaa gcc gac tcg ggg ggg cat acc gac aac agg      576
Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg
            180                 185                 190

cct att cac gtc att ctt cct cta atc atc aac cta cgc aac cgt ttg      624
Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu
        195                 200                 205

cat aaa gag ctt gac tac cct tcg cat ctc cgg gta cgt gtg ggt gct      672
His Lys Glu Leu Asp Tyr Pro Ser His Leu Arg Val Arg Val Gly Ala
    210                 215                 220

ggt ggt ggt att gga tgt cct caa gcc gct ctt gca gca ttt caa atg      720
Gly Gly Gly Ile Gly Cys Pro Gln Ala Ala Leu Ala Ala Phe Gln Met
225                 230                 235                 240

ggg gca gcg ttt tta atc act gga acg gtg aac cag ctt gct cgt gaa      768
Gly Ala Ala Phe Leu Ile Thr Gly Thr Val Asn Gln Leu Ala Arg Glu
                245                 250                 255

agt ggc act tgt gac aac gtc cgg tta cag ctc tca aag gcc acg tat      816
Ser Gly Thr Cys Asp Asn Val Arg Leu Gln Leu Ser Lys Ala Thr Tyr
            260                 265                 270

agc gac gtg tgt atg gct cct gct gcc gat atg ttt gac caa ggc gtg      864
Ser Asp Val Cys Met Ala Pro Ala Ala Asp Met Phe Asp Gln Gly Val
        275                 280                 285

gag ctg caa gta ttg aag aaa ggc acg ctg ttc cca agt cgt gct aag      912
Glu Leu Gln Val Leu Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys
    290                 295                 300

aag ctg tac gag ctg ttc tgc aag tat gac tcg ttt gag gca atg ccg      960
Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro
305                 310                 315                 320

gct gaa gaa ttg caa cgg gtt gaa aag cgg att ttt caa aag tcg ctt     1008
Ala Glu Glu Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu
                325                 330                 335

gct gaa gtt tgg cag gag acc agt gac ttt tac att cat cgt atc aag     1056
Ala Glu Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys
            340                 345                 350

aac cct gag aaa atc aat cgt gct gca agc gat ggc aaa ctg aaa atg     1104
```

```
Asn Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met
        355                 360                 365

tcg ctt tgc ttt cgc tgg tac ctt ggg ctt tcc tca ttt tgg gcc aac     1152
Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala Asn
    370                 375                 380

tct ggg gca caa gat cgc gtc atg gac tat caa att tgg tgt ggc cct     1200
Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys Gly Pro
385                 390                 395                 400

gct att ggc gct ttc aat gat ttt acc aag ggc acg tac ctt gac gtg     1248
Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr Leu Asp Val
                405                 410                 415

act gtt gca aag agt tac cct tgt gtg gca cag atc aat ttg caa att     1296
Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile Asn Leu Gln Ile
            420                 425                 430

ttg caa gga gct gcg tat ctg aaa cgc ctt ggt gtc att cgt ttt gac     1344
Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly Val Ile Arg Phe Asp
        435                 440                 445

cgc atg ctg ctg cag gcc gtc gat atc gac gat cct gta ttt act tac     1392
Arg Met Leu Leu Gln Ala Val Asp Ile Asp Asp Pro Val Phe Thr Tyr
    450                 455                 460

gtg ccg acc cag cca ctt                                             1410
Val Pro Thr Gln Pro Leu
465                 470
```

<210> 68
<211> 470
<212> PRT
<213> Thraustochytrium sp.

<400> 68

```
Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala Pro Leu Tyr
1               5                   10                  15

Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala
            20                  25                  30

Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
        35                  40                  45

Ile Ser Ile Val Arg Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro
    50                  55                  60

His Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
65                  70                  75                  80

Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val
                85                  90                  95

Val Glu Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr
            100                 105                 110

Arg Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala
```

115 120 125

His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe
130 135 140

Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala Ser Gly
145 150 155 160

Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val Pro Val Ala
165 170 175

Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg
180 185 190

Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu
195 200 205

His Lys Glu Leu Asp Tyr Pro Ser His Leu Arg Val Arg Val Gly Ala
210 215 220

Gly Gly Gly Ile Gly Cys Pro Gln Ala Ala Leu Ala Ala Phe Gln Met
225 230 235 240

Gly Ala Ala Phe Leu Ile Thr Gly Thr Val Asn Gln Leu Ala Arg Glu
245 250 255

Ser Gly Thr Cys Asp Asn Val Arg Leu Gln Leu Ser Lys Ala Thr Tyr
260 265 270

Ser Asp Val Cys Met Ala Pro Ala Ala Asp Met Phe Asp Gln Gly Val
275 280 285

Glu Leu Gln Val Leu Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys
290 295 300

Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro
305 310 315 320

Ala Glu Glu Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu
325 330 335

Ala Glu Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys
340 345 350

Asn Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met
355 360 365

Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala Asn
370 375 380

Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys Gly Pro

385 390 395 400

Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr Leu Asp Val
405 410 415

Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile Asn Leu Gln Ile
420 425 430

Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly Val Ile Arg Phe Asp
435 440 445

Arg Met Leu Leu Gln Ala Val Asp Ile Asp Asp Pro Val Phe Thr Tyr
450 455 460

Val Pro Thr Gln Pro Leu
465 470

<210> 69
<211> 6180
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 69
atggccgctc gcaacgtgag cgccgcgcat gagatgcacg atgaaaagcg catcgccgtc 60

gtcggcatgg ccgtccagta cgccggatgc aaaaccaagg acgagttctg ggaggtgctc 120

atgaacggca aggtcgagtc caaggtgatc agcgacaaac gactcggctc caactaccgc 180

gccgagcact acaaagcaga gcgcagcaag tatgccgaca ccttttgcaa cgaaacgtac 240

ggcacccttg acgagaacga gatcgacaac gagcacgaac tcctcctcaa cctcgccaag 300

caggcactcg cagagacatc cgtcaaagac tcgacacgct gcggcatcgt cagcggctgc 360

ctctcgttcc ccatggacaa cctccagggt gaactcctca acgtgtacca aaaccatgtc 420

gagaaaaagc tcggggcccg cgtcttcaag gacgcctccc attggtccga acgcgagcag 480

tccaacaaac ccgaggccgg tgaccgccgc atcttcatgg acccggcctc cttcgtcgcc 540

gaagaactca acctcggcgc ccttcactac tccgtcgacg cagcatgcgc cacggcgctc 600

tacgtgctcc gcctcgcgca ggatcatctc gtctccggcg ccgccgacgt catgctctgc 660

ggtgccacct gcctgccgga gccctttttc atcctttcgg gcttttccac cttccaggcc 720

atgcccgtcg gcacgggcca gaacgtgtcc atgccgctgc acaaggacag ccagggcctc 780

accccgggtg agggcggctc catcatggtc ctcaagcgtc tcgatgatgc catccgcgac 840

ggcgaccaca tctacggcac ccttctcggc gccaatgtca gcaactccgg cacaggtctg 900

cccctcaagc cccttctccc cagcgagaaa aagtgcctca tggacaccta cacgcgcatt 960

aacgtgcacc cgcacaagat tcagtacgtc gagtgccacg ccaccggcac gccccagggt 1020

gatcgtgtgg aaatcgacgc cgtcaaggcc tgctttgaag gcaaggtccc ccgtttcggt 1080

accacaaagg gcaactttgg acacaccctc gtcgcagccg gctttgccgg tatgtgcaag 1140

gtcctcctct ccatgaagca tggcatcatc ccgcccaccc cgggtatcga tgacgagacc 1200

aagatggacc ctctcgtcgt ctccggtgag gccatcccat ggccagagac caacggcgag 1260

cccaagcgcg ccggtctctc ggcctttggc tttggtggca ccaacgccca tgccgtcttt 1320

gaggagcatg acccctccaa cgccgcctgc acgggccacg actccatttc tgcgctctcg 1380

gcccgctgcg gcggtgaaag caacatgcgc atcgccatca ctggtatgga cgccaccttt 1440

ggcgctctca agggactcga cgccttcgag cgcgccattt acaccggcgc tcacggtgcc 1500

atcccactcc cagaaaagcg ctggcgcttt ctcggcaagg acaaggactt tcttgacctc 1560

tgcggcgtca aggccacccc gcacggctgc tacattgaag atgttgaggt cgacttccag 1620

cgcctccgca cgcccatgac ccctgaagac atgctcctcc ctcagcagct tctggccgtc 1680

accaccattg accgcgccat cctcgactcg ggaatgaaaa agggtggcaa tgtcgccgtc 1740

tttgtcggcc tcggcaccga cctcgagctc taccgtcacc gtgctcgcgt cgctctcaag 1800

gagcgcgtcc gccctgaagc ctccaagaag ctcaatgaca tgatgcagta cattaacgac 1860

tgcggcacat ccacatcgta cacctcgtac attggcaacc tcgtcgccac gcgcgtctcg 1920

tcgcagtggg gcttcacggg cccctccttt acgatcaccg agggcaacaa ctccgtctac 1980

cgctgcgccg agctcggcaa gtacctcctc gagaccggcg aggtcgatgg cgtcgtcgtt 2040

gcgggtgtcg atctctgcgg cagtgccgaa aacctttacg tcaagtctcg ccgcttcaag 2100

gtgtccacct ccgatacccc gcgcgccagc tttgacgccg ccgccgatgg ctactttgtc 2160

ggcgagggct gcggtgcctt tgtgctcaag cgtgagacta gctgcaccaa ggacgaccgt 2220

atctacgctt gcatggatgc catcgtccct ggcaacgtcc ctagcgcctg cttgcgcgag 2280

gccctcgacc aggcgcgcgt caagccgggc gatatcgaga tgctcgagct cagcgccgac 2340

tccgcccgcc acctcaagga cccgtccgtc ctgcccaagg agctcactgc cgaggaggaa 2400

atcggcggcc ttcagacgat ccttcgtgac gatgacaagc tcccgcgcaa cgtcgcaacg 2460

ggcagtgtca aggccaccgt cggtgacacc ggttatgcct ctggtgctgc cagcctcatc 2520

aaggctgcgc tttgcatcta caaccgctac ctgcccagca acggcgacga ctgggatgaa 2580

cccgcccctg aggcgccctg ggacagcacc ctctttgcgt gccagacctc gcgcgcttgg 2640

ctcaagaacc ctggcgagcg tcgctatgcg gccgtctcgg gcgtctccga gacgcgctcg 2700

tgctattccg tgctcctctc cgaagccgag ggccactacg agcgcgagaa ccgcatctcg 2760

ctcgacgagg aggcgcccaa gctcattgtg cttcgcgccg actcccacga ggagatcctt 2820

ggtcgcctcg acaagatccg cgagcgcttc ttgcagccca cgggcgccgc cccgcgcgag 2880

tccgagctca aggcgcaggc ccgccgcatc ttcctcgagc tcctcggcga gacccttgcc 2940

caggatgccg cttcttcagg ctcgcaaaag cccctcgctc tcagcctcgt ctccacgccc 3000

tccaagctcc agcgcgaggt cgagctcgcg gccaagggta tcccgcgctg cctcaagatg 3060

cgccgcgatt ggagctcccc tgctggcagc cgctacgcgc ctgagccgct cgccagcgac 3120

cgcgtcgcct tcatgtacgg cgaaggtcgc agcccttact acggcatcac ccaagacatt 3180

caccgcattt ggcccgaact ccacgaggtc atcaacgaaa agacgaaccg tctctgggcc 3240

gaaggcgacc gctgggtcat gccgcgcgcc agcttcaagt cggagctcga gagccagcag 3300

caagagtttg atcgcaacat gattgaaatg ttccgtcttg gaatcctcac ctcaattgcc 3360

ttcaccaatc tggcgcgcga cgttctcaac atcacgccca aggccgcctt tggcctcagt 3420

cttggcgaga tttccatgat ttttgccttt tccaagaaga acggtctcat ctccgaccag 3480

ctcaccaagg atcttcgcga gtccgacgtg tggaacaagg ctctggccgt tgaatttaat 3540

gcgctgcgcg aggcctgggg cattccacag agtgtcccca aggacgagtt ctggcaaggc 3600

tacattgtgc gcggcaccaa gcaggatatc gaggcggcca tcgccccgga cagcaagtac 3660

gtgcgcctca ccatcatcaa tgatgccaac accgccctca ttagcggcaa gcccgacgcc 3720

tgcaaggctg cgatcgcgcg tctcggtggc aacattcctg cgcttcccgt gacccagggc 3780

atgtgcggcc actgccccga ggtgggacct tataccaagg atatcgccaa gatccatgcc 3840

aaccttgagt tccccgttgt cgacggcctt gacctctgga ccacaatcaa ccagaagcgc 3900

ctcgtgccac gcgccacggg cgccaaggac gaatgggccc cttcttcctt tggcgagtac 3960

gccggccagc tctacgagaa gcaggctaac ttcccccaaa tcgtcgagac catttacaag 4020

caaaactacg acgtctttgt cgaggttggg cccaacaacc accgtagcac cgcagtgcgc 4080

accacgcttg gtccccagcg caaccacctt gctggcgcca tcgacaagca gaacgaggat 4140

gcttggacga ccatcgtcaa gcttgtggct tcgctcaagg cccaccttgt tcctggcgtc 4200

acgatctcgc cgctgtacca ctccaagctt gtggcggagg ctgaggcttg ctacgctgcg 4260

ctctgcaagg gtgaaaagcc caagaagaac aagtttgtgc gcaagattca gctcaacggt 4320

cgcttcaaca gcaaggcgga ccccatctcc tcggccgatc ttgccagctt tccgcctgcg 4380

gaccctgcca ttgaagccgc catctcgagc cgcatcatga agccggttgc tccgaagttc 4440

tacgcgcgtc tcaacattga cgagcaggac gagacccgtg atccgatcct caacaaggac 4500

aacgcgccgt cttccagctc tagctcctct tccagctctt ccagctcttc cagcccgtcg 4560

ccagctccgt ccgccccagt gcaaaagaag gctgctccgg ccgcggagac caaggctgtt 4620

gcttcggctg acgcacttcg cagtgccctg ctcgatctcg acagtatgct tgcgctgagc 4680

tctgccagtg cctccggcaa ccttgttgag actgcgccta gcgacgcctc ggtcattgtg 4740

ccgccctgca acattgcgga tctcggcagc cgcgccttca tgaaaacgta cggtgtttcg 4800

gcgcctctgt acacgggcgc catggccaag ggcattgcct ctgcggacct cgtcattgcc 4860

gccggccgcc agggcatcct tgcgtccttt ggcgccggcg gacttcccat gcaggttgtg 4920

cgtgagtcca tcgaaaagat tcaggccgcc ctgcccaatg gcccgtacgc tgtcaacctt 4980

atccattctc cctttgacag caacctcgaa aagggcaatg tcgatctctt cctcgagaag 5040

ggtgtcacct ttgtcgaggc ctcggccttt atgacgctca ccccgcaggt cgtgcggtac 5100

cgcgcggctg gcctcacgcg caacgccgac ggctcggtca acatccgcaa ccgtatcatt 5160

ggcaaggtct cgcgcaccga gctcgccgag atgttcatgc gtcctgcgcc cgagcacctt 5220

cttcagaagc tcattgcttc cggcgagatc aaccaggagc aggccgagct cgcccgccgt 5280

gttcccgtcg ctgacgacat cgcggtcgaa gctgactcgg gtggccacac cgacaaccgc 5340

cccatccacg tcattctgcc cctcatcatc aaccttcgcg accgccttca ccgcgagtgc 5400

ggctacccgg ccaaccttcg cgtccgtgtg ggcgccggcg gtggcattgg gtgcccccag 5460

gcggcgctgg ccaccttcaa catgggtgcc tcctttattg tcaccggcac cgtgaaccag 5520

gtcgccaagc agtcgggcac gtgcgacaat gtgcgcaagc agctcgcgaa ggccacttac 5580

tcggacgtat gcatggcccc ggctgccgac atgttcgagg aaggcgtcaa gcttcaggtc 5640

ctcaagaagg gaaccatgtt tccctcgcgc gccaacaagc tctacgagct cttttgcaag 5700

tacgactcgt tcgagtccat gccccccgca gagcttgcgc gcgtcgagaa gcgcatcttc 5760

agccgcgcgc tcgaagaggt ctgggacgag accaaaaact tttacattaa ccgtcttcac 5820

aacccggaga agatccagcg cgccgagcgc gaccccaagc tcaagatgtc gctgtgcttt 5880

cgctggtacc tgagcctggc gagccgctgg gccaacactg gagcttccga tcgcgtcatg 5940

gactaccagg tctggtgcgg tcctgccatt ggttccttca acgatttcat caagggaact 6000

taccttgatc cggccgtcgc aaacgagtac ccgtgcgtcg ttcagattaa caagcagatc 6060

cttcgtggag cgtgcttctt gcgccgtctc gaaattctgc gcaacgcacg cctttccgat 6120

ggcgctgccg ctcttgtggc cagcatcgat gacacatacg tcccggccga gaagctgtaa 6180

<210> 70
<211> 8394
<212> DNA
<213> Artificial

<220>
<223> synthetic

<220>
<221> misc_feature
<222> (227)..(227)

<220>
<221> misc_feature
<222> (227)..(227)
<223> n is a, c, g, or t

<400> 70
ccgcgaatca agaaggtagg cgcgctgcga ggcgcggcgg cggagcggag cgagggagag 60

ggagagggag agagagggag ggagacgtcg ccgcggcggg gcctggcctg gcctggtttg 120

gcttggtcag cgcggccttg tccgagcgtg cagctggagt tgggtggatt catttggatt 180

ttcttttgtt tttgtttttc tctctttccc ggaaagtgtt ggccggncgg tgttctttgt 240

tttgatttct tcaaaagttt tggtggttgg ttctctctct tggctctctg tcaggcggtc 300

cggtccacgc cccggcctct cctctcctct cctctcctct cctctccgtg cgtatacgta 360

cgtacgtttg tatacgtaca tacatcccgc ccgccgtgcc ggcgagggtt tgctcagcct 420

ggagcaatgc gatgcgatgc gatgcgatgc gacgcgacgc gacgcgagtc actggttcgc 480

gctgtggctg tggcttgctt gcttacttgc tttcgagctc tcccgctttc ttctttcctt 540

ctcacgccac caccaacgaa agaagatcgg ccccggcacg ccgctgagaa gggctggcgg 600

cgatgacggc acgcgcgccc gctgccacgt tggcgctcgc tgctgctgct gctgctgctg 660

ctgctgctgc tgctgctgct gctgctgctt ctgcgcgcag gctttgccac gaggccggcg 720

tgctggccgc tgccgcttcc agtccgcgtg gagagatcga atgagagata aactggatgg 780

attcatcgag ggatgaatga acgatggttg gatgcctttt tccttttca ggtccacagc 840

gggaagcagg agcgcgtgaa tctgccgcca tccgcatacg tctgcatcgc atcgcatcgc 900

atgcacgcat cgctcgccgg gagccacaga cgggcgacag ggcggccagc cagccaggca 960

gccagccagg caggcaccag agggccagag agcgcgcctc acgcacgcgc cgcagtgcgc 1020

gcatcgctcg cagtgcagac cttgattccc cgcgcggatc tccgcgagcc cgaaacgaag 1080

agcgccgtac gggcccatcc tagcgtcgcc tcgcaccgca tcgcatcgca tcgcgttccc 1140

tagagagtag tactcgacga aggcaccatt tccgcgctcc tcttcggcgc gatcgaggcc 1200

cccggcgccg cgacgatcgc ggcggccgcg gcgctggcgg cggccctggc gctcgcgctg 1260

gcggccgccg cgggcgtctg gccctggcgc gcgcgggcgc cgcaggagga gcggcagcgg 1320

ctgctcgccg ccagagaagg agcgcgccgg gcccggggag ggacggggag gagaaggaga 1380

aggcgcgcaa ggcggccccg aaagagaaga ccctggactt gaacgcgaag aagaagaaga 1440

aggagaagaa gttgaagaag aagaagaaga aggagaggaa gttgaagaag acgaggagca 1500

ggcgcgttcc aaggcgcgtt ctcttccgga ggcgcgttcc agctgcggcg gcggggcggg 1560

ctgcggggcg ggcgcgggcg cgggtgcggg cagaggggac gcgcgcgcgg aggcggaggg 1620

ggccgagcgg gagcccctgc tgctgcgggg cgcccgggcc gcaggtgtgg cgcgcgcgac 1680

gacggaggcg acgacgccag cggccgcgac gacaaggccg gcggcgtcgg cgggcggaag 1740

gccccgcgcg gagcaggggc gggagcagga caaggcgcag gagcaggagc agggccggga 1800

gcgggagcgg gagcgggcgg cggagcccga ggcagaaccc aatcgagatc cagagcgagc 1860

agaggccggc cgcgagcccg agcccgcgcc gcagatcact agtaccgctg cggaatcaca 1920

gcagcagcag cagcagcagc agcagcagca gcagcagcag cagccacgag agggagataa 1980

agaaaaagcg gcagagacga tgggtccccg tgttgcttcc ggtaaggttc cggcctggga 2040

gatgtcgaag tcggagctct gcgatgaccg cacggtcgtg tttgattacg aggagctcct 2100

tgagttcgcc gagggtgaca tctccaaggt gttcggtccg gagtttaaag tcgttgacgg 2160

ctttcgtcgc cgcgtccgcc tccccgcgcg tgagtacctc cttgtcaccc gcgtcacgct 2220

tatggacgcc gaggtgggta actttcgcgt gggcgctcgc atggttactg agtacgacgt 2280

tccggtcaac ggcgagctgt ctgagggcgg cgacgtgccc tgggccgttc ttgtcgaggc 2340

cggccagtgc gacctcctcc tcatctcgta catgggcatc gacttccagt gcaagggaga 2400

gcgcgtctac cgcctcctca acaccaccct tacgttcttc ggcgtcgcca aggagggcga 2460

```
gactctcgtg tacgacatcc gcgtcactgg tttcgccaag cgccctgacg gcgacatctc   2520
gatgttcttc tttgagtacg actgctactg caacggtaag ctcctcatcg agatgcgcga   2580
tggttctgcc ggctttttca ctgatgagga gctcgctgcc ggcaagggcg tcgtcgtcac   2640
gcgtgcccag caaaacatgc gcgacaagat cgtccgtcag tccatcgagc cgtttgcgct   2700
cgccgcctgc actcacaaga cgaccctcaa cgagtccgac atgcagtccc tcgtcgagcg   2760
caactgggcg aacgtgttcg gcaccagcaa caagatggcc gagctcaact acaagatttg   2820
cgcccgcaag atgctcatga tcgaccgcgt gacccacatc gaccatcatg gcggcgctta   2880
cggcctcggc ctccttgtcg gcgaaaagat cctcgatcgt aaccattggt acttcccctg   2940
ccactttgtg aatgaccaag ttatggccgg ctcccttgtc tctgacggct gttcgcagct   3000
gcttaagctt tacatgatct ggctcggcct ccacctcaag atggaggagt tcgattttct   3060
ccccgtcagc ggccacaaga acaaggtccg ctgccgcgga cagatttcgc cccacaaggg   3120
taagctcgtc tacgtgatgg agattaagaa gatgggctac gaccaggctt cgggcagccc   3180
ctacgccatt gcggacgttg atatcattga tgtcaacgag gagctcggcc agtcgtttga   3240
catcaacgat ctcgcgtcgt acggtaaggg cgacctctcg aagaagattg ttgtcgattt   3300
caaaggtatc gctctccagc tgaaaggccg cgccttctcc cgcatgtcct cctcgtcctc   3360
cctcaacgag ggctggcagt gcgtgcccaa gccgtcgcag cgcatggagc acgagcaacc   3420
ccctgcccat tgcctcgcct ccgatccgga ggcccctagc acggtgacgt ggcaccccat   3480
gagcaagctc cccggcaacc ccacccccct cttttctccc tcgtcgtatc cgccccgtgc   3540
tatctgcttc atccctttc cgggcaaccc gctcgacaac aactgcaagg ccggcgagat   3600
gcccctcaac tggtacaaca tgagcgaatt catgtgcggt aaggtcagca actgcctcgg   3660
cccggagttc gcccgctttg acaagagcaa cacctcccgc tcccccgcct tcgaccttgc   3720
gctcgtcacg cgcgtcgtcg aggtcaccaa catggagcac ggcaagttcc tcaacgtcga   3780
ctgtaaccct tcgaagggca cgatggtggg agagttcgac tgcccgcagg acgcctggtt   3840
ctttgacggc tcctgcaatg acggtcatat gccgtactcc atcatcatgg agatcggcct   3900
tcagacgtct ggtgtcctca cgtcggtgct caaggctcct ctcacgatgg acaaggacga   3960
catcctcttt cgcaaccttg atgcctcggc tgagatggtg cgccctgacg tcgatgtccg   4020
cggcaagacc atccgcaacg tgaccaagtg taccggctac gccatgcttg gtaagatggg   4080
tatccatcgc tttacctttg agctctccgt ggacggcgtc gtcttttaca agggcagcac   4140
cagcttcggc tggtttacgc cggaagtgtt tgcccagcag gccggtctgg acaacggcaa   4200
gaagaccgag ccctggtgca agaccaacaa cacctccgtc cgccgcgtcg agatcgccag   4260
cgctaagggt aaggagcagc tcacggagaa gctccctgat gctaccaacg cccaggttct   4320
ccgccgctcc gagcagtgcg agtacctcga ctacctcaac atcgcccccg actccggcct   4380
tcacggcaag ggttacgcgc acggccacaa ggacgtcaac ccccaggact ggtttttctc   4440
gtgccacttc tggttcgacc cggtgatgcc cggctcgctt ggcattgagt ccatgttcca   4500
```

gctcatcgaa gcctttgctg tcgaccagaa tatccctggc gagtacaatg tttcgaaccc 4560

gacctttgcc catgcccccg gtaagaccgc ttggaagtat cgcggccagc tcactcccaa 4620

gaaccgcgct atggactgcg aggtccacat cgtcagcatc acggcgagcc cggagaacgg 4680

cggctacgtc gacatcgtcg ccgacggcgc cctttgggtc gatggcctcc gcgtttacga 4740

ggccaaggag ctccgtgtcc gcgtcgtcag cgccaagccg caagctatcc cggatgtcca 4800

gcagcagcct ccctcggcta aggccgaccc tggcaagacc ggtgtcgccc tcagccccac 4860

ccagctccgc gatgtcctcc tggaggtcga caaccctctg tacctcggtg tcgagaactc 4920

caaccttgtc cagttcgaaa gcaagccggc gacgtcctcc cgcatcgtgt ccattaagcc 4980

gtgttctatc tccgacctcg gcgacaagtc tttcatggag acttacaacg tgtctgcgcc 5040

tctttacacc ggcgcgatgg ccaagggtat tgccagcgct gacctcgtca tcgccgctgg 5100

caagcgtaag atcctcggtt cgttcggtgc cggtggcctc ccgatctcca tcgttcgcga 5160

ggcccttgag aagatccagc agcacctccc ccacggccct tacgccgtca accttatcca 5220

ctccccgttc gactcgaacc tggagaaggg caatgtcgac ctttttcttg aaatgggtgt 5280

caccgtcgtc gagtgctcgg ctttcatgga gctcacggct caggtcgttc gctaccgcgc 5340

ttccggcctc tccaagagcg cggacggcag catccgcatc gcgcaccgca tcatcggcaa 5400

ggtgagccgc actgagcttg ccgagatgtt catccgtcct gcccctcagc accttctcca 5460

gaagctggtc gccagcggcg agctcaccgc cgaacaggcc gagcttgcca cgcaagtccc 5520

ggtcgccgat gacatcgcgg tcgaggccga ttcgggcggt cacaccgaca accgcccgat 5580

ccacgtgatt ctccccctta ttatcaacct ccgcaaccgc cttcacaagg agctcgacta 5640

cccctcgcac ctgcgcgtgc gcgtgggagc cggcggtggt atcggctgcc ctcaggcggc 5700

cctggcggcc tttcagatgg gcgctgcttt cctgattacc ggcaccgtca accagctcgc 5760

ccgtgagtcg ggcacctgcg acaacgtccg ccttcagctc tccaaggcca cctacagcga 5820

cgtgtgcatg gcccctgccg cggacatgtt tgaccagggc gtggagcttc aggtgctcaa 5880

gaagggcacg ctgttcccca gccgcgccaa gaagctctac gagctcttct gcaagtacga 5940

ctccttcgag gccatgccgg ccgaggagct tcagcgtgtc gagaagcgca tctttcagaa 6000

gtcgctggcc gaggtttggc aggagacttc ggatttctac attcaccgca ttaagaaccc 6060

tgagaagatc aatcgcgctg ccagcgacgg taaactcaag atgtcgcttt gcttccgctg 6120

gtatcttggc ctctcgtcgt tctgggccaa ctccggtgcc caggaccgcg tcatggacta 6180

ccagatttgg tgcggtcccg cgatcggagc cttcaacgat tttaccaagg gcacctacct 6240

cgacgtcacg gttgccaagt cttacccgtg cgttgcccag atcaacctcc aaatccttca 6300

gggtgcggct tacctcaagc gcctcggtgt gatccgcttc gaccgcatgc tcctccaggc 6360

ggtcgacatc gatgacccgg ttttcacgta cgtccccacc cagcccctct aagaaagtga 6420

accttgtcct aacccgacag cgaatggcgg gaggggggcgg gctaaaagat cgtattacat 6480

agtatttttc ccctactctt tgtgtttgtc ttttttttttt tttgaacgca ttcaagccac 6540

ttgtctbggt ttacttgttt gttygcttgc ttgcttgctt gcttgcctgc ttcttggtca 6600

gacggcccaa aaaagggaaa aaattcattc atggcacaga taagaaaaag aaaargtttg 6660

tcgaccaccg tcatcagaaa gcaagagaag agaaacactc gcgctcacat tctcgctcgc 6720

gtaagaatct tagccacgca tacgaagtaa tttgtccatc tggcgaatct ttacatgagc 6780

gttttcaagc tggagcgtga gatcatacct ttcttgatcg taatgttcca accttgcata 6840

ggcctcgttg cgatccgcta gcaatgcgtc gtactcccgt tgcarctgcg ccatcgcctc 6900

attgtgacgt gagttcagat tcttctcgag accttcgagc gctgctaatt tcgcctgacg 6960

ctccttcttt tgtgcttcca tgacacgccg cttcaccgtg cgttccactt cttcctcaga 7020

catgcccttg gctgcctcga cctgctcggt aagcttcgtc gtaatctcct cgatctcgga 7080

attcttcttg ccctccatcc actcggcacc atacttggca gcctgttcaa cacgctcatt 7140

gaaaaacttt tcattctctt ccagctccgc aacccgcgct cgaagctcat tcacttccgc 7200

caccacggct tcggcatcga gcgccgaatc agtcgccgaa ctttccgaaa gatacaccac 7260

ggcccctccg ctgctgctgc gcagcgtcat catcagtcgc gtgttatckt cgcgcagatt 7320

ctccacctgc tccgtaagca gcttcacggt ggcctcttga ttctgagggc tcacgtcgtg 7380

gattagcgct tgcagctctt gcagctccgt cagcttggaa gagctcgtaa tcatggcttt 7440

gcacttgtcc agacgtcgca gagcgttcga gagccgcttc gcgttatctg ccatggacgc 7500

ttctgcgctc gcggcctccc tgacgacagt ctcttgcagt ttcactagat catgtccaat 7560

cagcttgcgg tgcagctctc caatcacgtt ctgcatcttg tttgtgtgtc cgggccgcgc 7620

ctcgtcttgc gatttgcgaa tttcctcctc gagctcgcgt tcgagctcca rggcgccttt 7680

aagtagctcg aagtcagccg ccgttagccc cagctccgtc gccgcgttca gacagtcggt 7740

tagcttgatt cgattccgct tttccatggc aagtttaaga tcctggccca gctgcacctc 7800

ctgcgccttg cgcatcatgc gcggttccgc ctggcgcaaa agcttcgagt cgtatcctgc 7860

ctgccatgcc agcgcaatgg cacgcacgag cgacttgagt tgccmactat tcatcgccga 7920

gatgagcagc attttgatct gcatgaacac ctcgtcagag tcgtcatcct ctgcctcctc 7980

cagctctgcg ggcgagcgac gctctccttg cagatgaagc gagggccgca ggcctccgaa 8040

gagcacctct tgcgcgagat cctcctccgt cgtcgccctc cgcaggatyg cggycgtgtc 8100

cgccatcttg ccgccacagc agcttttgst cgctctgcac cttcaatttc tggtgccgct 8160

ggtgccgctg gtgccgcttg tgctggtgct ggtgctggtg ctggtgctgg tgccttgtgc 8220

tggtgctgcc acagacaccg ccgctcctgc tgctgctctt ccggcccccc cgccgccgcc 8280

gcgagccccc gccgcgcgcc gtgcctgggc tctccgcgct ctccgcgggc tcctcggcct 8340

cggcctcgcc gtccgcgacg acgtctgcgc ggccgatggt gcggatctgc tcta 8394

<210> 71
<211> 12495
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 71
cgactgcgag ggctcttgcg tcatccgcct gatgcggccg ccggccggaa gaggccgccc 60

cgcagcctcc ccggtgcggg gccgcgccgg gatgcctcgc ctgcgtcccc agcgatgcga 120

ggcggcgacg tctgtgctgc cgtcaggtcc tgtgcgaggt cccactgtta atattaagta 180

cgtaccgcac aggacccgga gtagcgcggc gcggggacga cctccgacgc ggccggaaga 240

cggcgcccac agcacgcgct cgggttcgcg ggcggcggcc gcgcgaagtg cgtaagcaga 300

cggcgcgcgg tccggctgcg ggggaagaag gtgtggaagg tgcgggggga acccgtggcc 360

gcctacgtgt gcctgcgacg gaggcgcggc ggcgcccggc acctaggaac tgtctaagag 420

ggcttacaac ggacagcccg tcacaggcgc agaggcggcg gcggcgggtc tttccttctc 480

tcccggcaga aagagaggga gggagagaga gagaaagcct gaatgtggct gcgatggaga 540

cttgggcagc gataatgact tggtgggctg aagattactc tcagcattgc tgcgcgctta 600

gcattcttgc ggcattggcc aggtcggcgt cgggtttcgc ttgcggcgac gaccgagggt 660

cgttggatcg cgcggcagat tagacaggca gctgggacgc acgcacgcac gcacgaacga 720

acaaacaaac gagagaacgt gcgtacgggc ctgctggccg cggatgcaga aaagtgaatg 780

gaattttcca ctacagccag cgagccctgt cagccagcta gctgcagctt gagcggagcg 840

gcggctgttg ctgcgtctct tgaacttgaa ttcgaatgga tgtatgtact tgcttgcctc 900

ttgatgctgc gttgacttgt ggatgatgtt caaagtggca ggcacaccac ctaaaatcaa 960

aaaatgtaaa ggcgcggcgc agaaccgtgc atctgccgag gagcgcgtgc accgctgccg 1020

tgcgtgcggt gcgtgcgtgc tcgcctgcat ccgcgtcgcc ggcaaggtct ctgtgctgta 1080

tctatcgctc gcattcgaat cgctcgcgcc ttctgctccg acgtgcccta aaacatagag 1140

tccatcatta ggctagtgat actctatgtg acttgcttcc ttcctggaga ggaagaatga 1200

ataccgtatg gtatcaggta tgaagaaaac caacctcgcg cctccggacc caccgccgcg 1260

cggccagcca gatccgcgcg gccagctcgt ccatgccagc ctcgggaggt atccatcccc 1320

tccgcgccgg cagctgtcgg tctctcgcca tcgcctgtgc cgctcgaaga agtttcagca 1380

acctctccgt cctgcgccat ggcaccatct tatcttctcc gcgtcccact cttctctccc 1440

aaagcctgca cctcaaaggc gaggccgggc caggccgcag ccgcgccacg agaagcccca 1500

tcacgccgcc tcgcacgcgg ccgcgcgcgt ctgcgcgaga cggcgctgtt attttctttt 1560

cttttctttt cttttcttct ctcctccccg ctctcctcgc ttcctgcgcg ccgctgcggg 1620

ggcgaaggaa gacaggggcg ctcgggagga agaggaggcg agggaaggaa ggagggcggc 1680

gacggcatcg ccagacgacg acggcggcgg cgggggacgc ggggcgcgcg tccccccgtg 1740

cgtggcagca tgatgcatgc gatgcgcgcg atgcgtgcgg cgcgaggagg cgaggcccga 1800

tgggcggagg cgagcggggt ggatacgggc ggcggcgggc gggagcggga ggcaggcaga 1860

ccccgaaccc cgagggtaaa actacacaaa cagacagaca gcctgcggct gcctggcaac 1920

aagactcacg cacgcacgca cgcaccgatc ggtcgagcga gcgagcgctt gatcgaacgg 1980

acgtgagaac cagcaaggaa gcagcagcag cagcagcagc agcaacagct cgagagagag 2040

aaaatgaagg atatggagga tcgccgcgtc gccattgtcg gtatgtcggc ccatctcccc 2100

tgcggcaccg acgtgaagga gtcgtggcag gccattcgtg atggtatcga ttgcctgtct 2160

gacctccccg ccgaccgcct cgacgtcacg gcgtactaca accccaacaa ggccactaag 2220

gataagattt actgcaagcg cggtggcttt attcccaact acgattttga tcctcgcgag 2280

tttggcctta acatgttcca gatggaggac tccgacgcca accagactct cactcttctg 2340

aaggtcaagc aggccctgga ggacgcttcc attgagccct ttaccaagga gaagaagaac 2400

atcggctgcg tcctcggaat cggcggtggc cagaaggcca gccacgagtt ttactctcgc 2460

ctcaactacg ttgtcgtcga gaaggtcctc cgcaagatgg gcctccctga tgccgacgtt 2520

gaggaggccg tggagaagta caaggccaat tttcctgagt ggcgcctcga cagcttccct 2580

ggcttcctcg gcaacgtcac tgctggccgc tgctccaaca cctttaacat ggagggtatg 2640

aactgcgtgg tggacgccgc ctgcgcctcc tccctcatcg ccattaaggt ggcggtggag 2700

gagctcctct tcggcgactg cgacaccatg attgccggtg ccacctgcac ggataactcg 2760

ctcggcatgt acatggcttt ttccaagacg ccggtgttca gcaccgaccc ctccgtccgc 2820

gcctacgatg aaaagactaa gggcatgctt atcggcgagg gctccgccat gttcgtgctc 2880

aagcgctacg ccgacgcggt ccgtgatggc gacactatcc acgccgttct ccgtagctgc 2940

tcgtccagca gcgacggcaa ggccgctgga atttacacgc ctacgatctc gggccaggag 3000

gaggccctcc gccgtgccta cgcccgcgct ggcgtctgcc cctcgaccat cggcctcgtt 3060

gagggtcacg gcaccggcac gccggtcggc gaccgcatcg agctcactgc cctccgcaat 3120

ctcttcgaca aggcttttgg ctccaagaag gaacaaatcg ctgtcggttc catcaagtct 3180

cagatcggcc acctcaagag cgtggccggc tttgctggcc tggtcaaggc ggtcctggcc 3240

ctcaagcaca agacgctccc tggctcgatt aacgtggacc agcctcccct cctttacgac 3300

ggcacgcaga tccaggactc gtccctgtac attaacaaga ctaaccgccc ctggtttacc 3360

caaaacaagc tccctcgccg tgcgggtgtc agctccttcg gttttggtgg cgccaactac 3420

cacgccgtcc tcgaagagtt tgagcccgag cacgagaagc cttaccgcct caacacggtc 3480

ggccaccccg tcctcctcta cgcccccagc gtcgaggccc tcaaggtcct ttgcaacgac 3540

cagctcgccg aactcaccat cgcgcttgag gaggccaaga cccacaagaa cgtggacaag 3600

gtttgcggtt acaagtttat cgacgagttc cagctccagg gtagctgccc tcctgagaac 3660

ccccgcgtcg gctttctcgc cactctcccg acgtccaaca tcatcgtcgc tctcaaggcg 3720

atcctcgctc agctcgatgc caagcctgac gccaagaagt gggacctccc gcacaagaag 3780

gctttcggtg ctaccttcgc ctcgtcgtcg gtcaagggct cggttgccgc gctgttcgcc 3840

ggccagggca ctcagtacct caacatgttt tccgatgtcg cgatgaactg gccccctttc 3900

cgcgattcga tcgtggccat ggaggaggcc cagaccgagg tgttcgaggg ccaggttgaa 3960

cccatttcga aggtcctctt tccgcgcgag cgttacgcct cggagtcgga gcagggcaac 4020

gagctgctct gcctcaccga gtactcgcag cccactacaa ttgccgccgc tgtcggcgcc 4080

tttgacatct tcaaggcggc tggctttaag cccgatatgg tcggaggcca ctcccttggc 4140

gagtttgctg ccctgtacgc cgctggctcg atctctcgcg atgacctcta caagctcgtc 4200

tgcaagcgtg ccaaggcgat ggcgaacgcg agcgatggag ccatggccgc tgtgatcggt 4260

cccgatgctc gcctcgtcac tcctcagaac tccgacgtct acgtcgccaa tttcaatagc 4320

gcgacgcagg tcgtcatcag cggcacggtg cagggcgtca aagaggagag caagctcctt 4380

attagcaagg gttttcgcgt cctgcctctg aagtgccagg gcgccttcca ctccccgctc 4440

atgggacctt ccgaggactc cttcaagtcg ctcgttgaga cttgtactat ctccccgcct 4500

aagaacgtca agttcttttg caacgtgagc ggcaaggagt cgcctaatcc gaagcagact 4560

ctcaagtctc acatgaccag ctccgtccag tttgaggagc agatccgcaa catgtacgat 4620

gccggtgccc gtgtctttct ggagtttggc cctcgccagg tgctcgccaa gctcatcgcc 4680

gagatgttcc cctcgtgcac cgccatctcg gtcaacccgg cctcctcggg cgactcggac 4740

gtccagctcc gcctcgcggc cgtcaagttt gcggtgtccg gcgctgccct ctcgacgttt 4800

gacccctggg agtaccgcaa gccccaggac ctccttatcc gcaagccccg caagaccgcg 4860

ctcgtcctct ctgctgccac gtacgtttcg cctaagaccc tcgccgagcg caagaaggcc 4920

atggaggaca tcaagctcgt gtcgatcacc cctcgcgact cgatggtttc gattggtaag 4980

atcgctcaag aggtccgcac tgctaagcag cccctcgaaa ccgagattcg ccgcctcaac 5040

aaggagctcg aacacctcaa gcgcgagctc gcggctgcca aagcttcggt caagtctgcc 5100

tccaagtcct ctaaggagcg cagcgtcctc tccaagcacc gcgccctcct ccagaacatc 5160

ctccaagact acgacgacct ccgcgtcgtc cccttcgccg tccgctcggt cgccgtggac 5220

aacaccgcgc cgtacgccga ccaggtcagc acccctgcct cggagcgtag cgcgtcgccc 5280

cttttcgaga agcgttcctc cgtttcgtcg gctcgcctgg ccgaggcgga ggccgccgtc 5340

ctctctgtgc tcgccgacaa gacgggctac gacagctcga tgatcgagat ggacatggac 5400

ctggagagcg agcttggcgt ggactccatc aagcgcgtcg agatcatgtc cgaggtccag 5460

actcttctct cggtcgaggt cagcgatgtt gacgctctca gccgcaccaa aacggtcggc 5520

gacgtgatcg aggctatgaa gctcgaactt ggcggacccc aaggccagac cctcaccgcc 5580

gagtcgatcc gtcagcctcc cgtgtcggag cccgccgtcc cgacgtcctc ctcttcctcc 5640

atcgctaacg tctcttctgc tcgcctcgcc gaggctgaag ctgctgttct ttctgtcctc 5700

gccgacaaga ccggctacga ctcgtccatg attgagatgg acatggacct cgaaagcgag 5760

ctcggcgtcg acagcattaa gcgcgtcgag atcatgtccg aggtgcagac ccttctcagc 5820

gtggaggttt cggacgtcga tgccctctcc cgcactaaga ccgtcggcga tgtcatcgag 5880

gccatgaagc tcgaactcgg cggtcctcag ggtcaaacgc tgaccgccga gtcgattcgc 5940

cagccccctg tttccgagcc ggccgtgccg acgtcgtcgt cgagctcgat cgccaacgtt 6000

tccagcgccc gcctcgccga ggccgaagcc gccgtcctct ccgtcctcgc ggacaagacc 6060

ggttacgact cctcgatgat tgaaatggac atggatcttg agtctgagct tggcgtcgat 6120

tccatcaagc gcgttgagat catgtcggaa gtgcaaactc tcctttctgt tgaagtctct 6180

gatgtcgatg ctctcagccg cacgaaaacc gtcggtgacg tcatcgaggc catgaagctc 6240

gaactcggtg gccctcaggg ccagaccctc accgccgagt ccatccgtca accgcccgtg 6300

tccgagcccg ccgtcccgac ctcctcgtcg agctcgattg cgaacgttct ctcggcgcgt 6360

ctcgcggagg ccgaggccgc ggtcctctcg gtcctcgccg acaagacggg ctacgactcc 6420

agcatgatcg agatggacat ggacctcgaa tctgagctcg gcgtcgacag cattaagcgc 6480

gtcgagatta tgagcgaggt ccagaccctc ctctccgtgg aggtttcgga cgtggacgcg 6540

ctctcgcgta ccaagaccgt tggcgacgtt attgaggcca tgaaactcga actcggtggc 6600

cctcagggac agaccctcac tgccgagtcg attcgccagc ctccggtttc tgagcctgct 6660

gttcctacgt cctcttcttc ttctattgct aatgtcagct cggcgcgcct cgcggaggcc 6720

gaggctgccg tgctctcggt tctcgcggac aagaccggct atgactcgtc catgattgag 6780

atggatatgg atcttgagtc cgagctgggc gtcgatagca tcaagcgtgt ggaaatcatg 6840

tccgaagtcc agaccctcct ctcggtcgag gtcagcgatg tcgacgccct ctcgcgcacc 6900

aagaccgtcg gcgatgtcat cgaggccatg aaacttgagc tcggcggacc tcagggacag 6960

accctcacgt ccgagccgat ccaccagccg cccgtgagcg agccggccgt cccgacttcg 7020

agctcctcgt ccatcgccaa tgtctctagc gctcgcctcg ccgaggccga ggccgccgtg 7080

ctgtccgtcc tcgccgataa gaccggatat gactccagca tgatcgagat ggacatggac 7140

cttgagtccg agctgggtgt tgattctatt aagcgtgttg agattatgtc ggaggtgcag 7200

acccttctct cggtcgaagt ctctgatgtc gatgccctct ctcgtaccaa gactgtcggc 7260

gacgtcatcg aggctatgaa gatggagctc ggcggacccc agggccagac tcttaccgcc 7320

gaatctatcc gccagccgcc tgtcagcgag cctgccgtcc ctacctcctc gtcctcgtcg 7380

attgccaacg tttcttctgc ccgtcttgcc gaggccgagg cggccgtcct ctcggtcctc 7440

gcggacaaga ccggttatga ttcgtcgatg attgagatgg acatggacct tgagtcggag 7500

ctcggagtcg acagcatcaa gcgcgtcgag atcatgtcgg aggtccaggc tctcctctcc 7560

gtcgaggtta gcgatgtcga cgccctctcg cgcaccaaga ccgttggcga tgttattgag 7620

gccatgaaaa tggagctcgg tggtccgcag ggccagaccc tcactgccga gtcgattcgc 7680

gagccgccgg tttccgagcc ggccgtgccc acgagcagct cgtcgagcat cgctaacgtc 7740

agctcggccc gccttgccga ggccgaggct gctgtcctct ccgtgctcgc cgataagact 7800

ggttatgact cgtccatgat tgagatggac atggacctgg agtccgagct cggcgtggac 7860

tcgatcaagc gcgttgagat catgtctgag gtgcagaccc tgctttccgt cgaggtgagc 7920

gacgtcgatg ccctgtcgcg taccaagact gtgggtgacg tcatcgaggc gatgaagctg 7980

gagctcggag agtcgagctc gattgagact cttaattgca ccgaggtcga gcacacgtcc 8040

tacaagagcg tcaaggccag cggatgcgag aacgtggaca cccgtttcgc caaagtggtc 8100

cagatcagcc tcccgtccaa gcttaagagc accgtcagcc atgaccgccc ggtcattgtt 8160

gtggatgacg gcaccccgct taccacggag ctctgcaaga ttctcggtgg caacattgtc 8220

gttctcagct accagggcaa gcctgccggc cctcgtggcg tcgaggttcc cgacctctcg 8280

gaggaggccc tcatccaggc gctcgccctc attcgctcga cctacggcgt ccccattggt 8340

ttcatctgcc agcaggtttc caacgtcagc accaaggccc agctctgctg ggcgctgctg 8400

gccgctaagc acctcaagaa ggacctcaat gccgtcctcc ctgactcccg ctcgtttttc 8460

gtgggcgttg tccgcctcaa cggcaagctg ggcacctttg agaacatttc cgactttagc 8520

aagttcgatc tcactaaggc gctcgactac ggccagcgcg gctccctcct cggcctgtgc 8580

aagagcctcg atcttgagtg ggagcaggtt ttctgccgcg gcatcgacct ggcctgcgac 8640

ctcatgcctc tccaagccgc gcgtatcctc cgcaacgagc tccagtgccc gaacatgcgc 8700

ctgcgcgagg tcggctacga tatctcgggc gctcgctaca ccatctccac cgacgacctt 8760

ctttgcggcc ctagcaaggc caaggtcgaa gcggcggacc tctttcttgt cacgggtggt 8820

gcccgcggta tcactcccca ctgcgtccgc gagatcgcct cgcgtagccc cggcacgacg 8880

tttgttctcg tcggccgctc cgagatgagc gacgagcccg actgggccgt gggtcactac 8940

aacaaggatc tcgaccagtc cacgatgaag cacctgaagg ccacgcatgc tgccggaggc 9000

gtgaagccga ctcccaaggc gcaccgcgcc cttgtcaacc gtgtcaccgg ctcgcgcgag 9060

gtccgcgaga gcctgcgcgc cattcaggag gcgggtgcga acgttgagta catcgcctgc 9120

gacgtcagcg acgagaacaa ggtgcgccag ctcgtccagc gcgtggagca gaagtacggt 9180

tgcgagatta cgggcatctg gcacgcctcg ggtgttctcc gcgacaagct cgtggagcaa 9240

aagaccacgg acgacttcga ggccgtgttt ggcactaagg tcacgggcct cgtgaacatc 9300

gtcagccagg tcaacatgtc caagctgcgc catttcatcc tcttctcctc gctcgcgggc 9360

tttcacggca acaagggcca gaccgactac gccatcgcca acgaggccct caacaagatc 9420

gcccacaccc tttccgcctt tctccccaag ctcaacgcga aggtcctcga ctttggcccg 9480

tgggtcggct ccggcatggt cacggagact cttgagaagc actttaaggc catgggcgtc 9540

cagaccatcc ccctcgaacc gggtgcccgc accgtggccc agatcatcct cgccagctct 9600

ccgccccagt ccctcctcgg taactggggc ttccccgcca cgaaacccct tcagcgctcc 9660

aacgtggtca ccggcaccct ctcgccggag gagatcgaat ttatcgccga tcacaagatc 9720

cagggccgca aggttctccc gatgatggcc gccatcggct tcatggcgtc gatcgccgag 9780

ggcctctacc ctggctacaa ccttcagggc gttgagaacg ctcagctttt tcagggcctc 9840

accatcaacc aggaaaccaa gttccagatc acgctcatcg aggagcacaa ctccgaggag 9900

aacctcgacg tcctcaccag cctcggtgtc atgctcgaaa gcggcaaggt gcttcccgcc 9960

taccgttgcg tcgtctgcct caacactacc cagcagcagc ccaagctcag ccccaagatc 10020

ctcaacctgg aggttgaccc cgcctgcgag gtcaacccct acgacggcaa gtccctcttt 10080

cacggcccgc tcctccaatt tgtccagcag gtcctccact cgtcgacgaa aggcctcgtg 10140

gcgaagtgcc gtgccctgcc gattaaggag gccatccgtg gcccgttcat caagcagacg 10200

ctccacgacc cgatcctcga cgacgtcatc tttcagctca tgctcgtgtg gtgccgcaac 10260

gccctcggct cggcctccct cccgaaccgc attgagaaga tgagctactt cggtaatgtg 10320

tcggagggca gcactttctt cgccagcgtt acccccgtcg gcccccgtgt cccgaaggat 10380

cccgtgatca agatgcagtt tctcctccag gacgagtccg gcaacacctt ttcgtccggc 10440

gagggcagcg tcgtcctcag cgatgagctc gtgttctaaa tctaaactag tgaatcgcga 10500

caagttgtct ttttgttact ctacgtactg ctctttctaa ttttacgacg tatgcttctg 10560

ctcttgacga cgacaaccat ggcaataaaa gtaaggcaag aaagtgagtg cgtgctcgct 10620

cacaaggtct aggccgaaac gaggcgcctt aggattggtt gtctccgtcg taagtcatga 10680

cggagcgtag gacaccgacg gcctgccgcg caaatatacg aaccgctgca caattctttt 10740

cgttgagaac aacctcgaac ggcctgcctt gctgagctgc cgacacggtg gaaggaagca 10800

tagcggccaa tcgaggggat gctattaatt aggcttggcg ctgctgccgc cgtgcctgaa 10860

gatgtgcgca agcgtggcaa caccgtccac cttgccatta aagtcatgca ctccaaatac 10920

ctcgaacaga gcctctcgca tgtttggcca agcaccttcg agagagtctg gactctcact 10980

ccttcctcac cgttgtcgcc attctcagcg agctcatgca atcaccagga cggagagacg 11040

gccagctgct tttctgcttt ccattgttat tagagaaacg cttctcgctc tcatcgtctt 11100

agtagacatt ccgatggctt cgttcgccaa tttgtcacct aagtaagcta gagtgttaag 11160

tctaaatgcc tttgacccgc gtacggcgtc acgtagatgc ctgtccttgc cagcaaacgc 11220

tagttcgcgg tgtgcgtaat ttggcccgca ttatgctggc tctcaaaatc aaccgccacc 11280

actcgcggct gcacgatgat tttcgtgcac tcatgacatg agaaatgtga tactcaaact 11340

agtatagacc tcctactcct actgctgctt ttctcgtcag agctgtctcc ggaaaagttg 11400

acaagttgtt ggccttcttc ttctctgcta gtaagtagat catcatggat gagacgatga 11460

tgataatgat gatgatgata gcgaagagct tttgggactc gagcttgagc tgatgtggct 11520

tattttatta ctcaaaaacc aatgcaccgt tgcgagctct cgagagtgga gcaaagttgc 11580

aacagcgcac acacaaagaa cacatagaaa tagacatcaa aaatagtact tttacaatcc 11640

agccgaggca gcgcgaaaag tcagcatctg cgaactgcca caaaagccgt agatcaagtt 11700

ttgacacaac gctcatgcac gtctacgtgc tcgcgcatat aggaaaacaa atcacatcga 11760

gagattccca ctcaggtcgt gcgtggcctt ttgggcatac tcttgccgcg cagcagccca 11820

caacgcctca tcttctccca taccaaggtc aaaggatggt tgtattttgt cgtgaagttg 11880

gggcggcatt ttctctaaaa gcattgatgc aatagatttc gggctgtagc gcgcggaaaa 11940

atgcatgagc agaagcgtat cgacgcgatc aaacaagcct cgatgaccga acgctgcaag 12000

ttggtcaaca tgaatatggc cgtggtcttc ggccttggaa atgtcgccat cgaaaaaggt 12060

cacctccacg cagagcgtct ttgcacggag aaagggagat tcagggtccc gttccaaagc 12120

agcagcgacc gtgtcgcccg tgtaagcaaa gtcgacaacc tcgtgtatga cgtagtcctc 12180

aggctttggt ggcgaggtct tgaagcgagt gcgctttgtc gcaacgggga gctccgcaag 12240

ttccggacgt aactccttttt tgatattaat gaccgcatat cctacagaag gcacacgatg 12300

ctcggttgca aggactgcaa cgcgcgtatt cccgttcaac tctacgctgt cgccggggtt 12360

gacgcccgtg atcttcatat tgaaggggtc ctgggtgtcg tcaacatctc ctcctccgtc 12420

gagggcttgc atcgctcgcc ctgcactgcg aaaatcctcc acagcctccg caggcatgta 12480

gacgcgcgtc tcctt 12495

<210> 72
<211> 8647
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 72
ctatactggt gctcttcggt gctgatttag ctcgacttgg gtccacacca ctgcccctct 60

actgtgaatg gatcaatgga cgcacgacgg gccgacgaaa gtgcgcgagt gaggtaacct 120

aagcaacggc ggtcttcaga ggggacgcac gccctccgtc gcagtcagtc cagacaggca 180

gaaaagcgtc ttagggacca cgcacgcacg cacgcacgca cgcacgcccg cacgcacgct 240

ccctccctcg cgtgcctatt tttttaggct tccttccgca cgggcctacc tctcgctccc 300

tcgcctcgcc gcaccaggcg gcagcagcga tacctgccgg tgccgcctcc gtcacgcgct 360

cagccgcagc tcagcccagc cgcgagctag ggtttgttcg tcctgaattg tttgatttga 420

tttgatttga tttgatccga tccgatccga tctgatctga tttgctttgc tttgctttgt 480

ctccctcccg gcgcggacca agcgtccgtc tgcgcgccgc agcttccctt cttctcccag 540

ccctccttct gctcccgcct ctcgcgcaag cacgcagctt cgccgccgca tccggtcggt 600

cggtcggtcg atcgacccgc ctgccgctgc tgctgtggcc gggcttttct ccatcggcga 660

ctctttcttc tccatacgtc ctactacgta catacatact gccggcttcc tcctcttcca 720

gcgcggcgac ggcggcaggc tgcgacgtcg tcgccgccgc gggcgccgcg cgcgccgccg 780

ccgccgcccg cgtcgcaggg cctcgtcgcc gccgccgctc cgctccgctc cgaggccgcg 840

agagggccgc ggcggcgcga tggatggatg gatggatgga tggatggatg gattttgttg 900

atcgatggcg gcgcatgggc ggagatgagc gaggacgagc gcgcgagcgc ggcagccgga 960

ttcgcagggc ctcgctcgcc tcgcgcccgc tgccgcgccc gccttgcgag cctgcgccgc 1020

gagcgagcga gcgagcgagc ggggctttct ttgtctcgcg cgccgcttgg cctcgtgtgt 1080

cttgtgcttg cgtagcgggc gccgcggtgg aagatggctc attcaatcga cccattcacg 1140

cacgcactcc ggcgcgcaga gaaggccgag gaggagcagc aagcaaacca aaagctctcg 1200

cgctcgcggt ctcgggctcg agcggtctcg gagagagagt cttgcggcga ccaccggcag 1260

cagcagcagc agcagcagcg ctgtcgagca cgagcacgag cacgagcacg agcacgagca 1320

ttcgagcaag aggacagaca cggttgtcag cgcctagctc gctcgataca gaaagaggcg 1380

ggttgggcgt aaaaaaaaag gagcacgcaa gccgccagcc agccagctag ctagccagcc 1440

tgcctgccaa aatgcagctc ccgcccgctc actccgctga tgagaaccgc attgccgtgg 1500

tcggcatggc cgtgaagtac gctggctgcg acaacaagga ggaattctgg aaaaccctca 1560

tgaacggttc gattaacacc aaaagcatta gcgcggctcg ccttggctcg aacaagcgcg 1620

acgagcacta cgtcccggag cgcagcaagt acgccgacac gttctgcaac gagcgttacg 1680

gctgcatcca gcagggcacg gataacgagc acgacctcct cctcggcctc gcccaggagg 1740

ccctcgccga tgctgccggc cgtatggaga agcagccgtc ggaggccttc gatcttgaga 1800

ataccggcat cgtgtcgggc tgcctgagct ttccgatgga taacctccag ggcgagctcc 1860

tcaacctcta ccagtcccac gtcgagaagc agcttccccc gtctgctctc gtggaggccg 1920

tcaagctctg gagcgagcgc cagaagtcca ctaaggccca cgcgggcgac aagcgccgct 1980

ttatcgatcc ggcgtctttc gtcgccgaca agctgaacct cggccccctc cactacgcca 2040

tcgacgccgc ctgcgcctcg gcgctctacg tgctgaagct cgcccaggac cacctcgtca 2100

gcggcgctgt cgatatgatg ctctgcggcg ccacttgttt cccggagccg ttcttcatcc 2160

tctcgggctt cagcaccttc caagccatgc cggtcggtgc ggacggcgtg tcgctccccc 2220

tccacaagac ctcggctggc ctcacccccg gtgagggcgg ctccatcatg gttctcaagc 2280

gcctcaagga cgccatccgc gacggcaacc acatctacgg cgtgctcctt gaggccaacc 2340

tctcgaacgc cggatgcggc ctcccgctct cgccgcacct cccttccgag gagtcgtgca 2400

ttcgcgacac ctatcgccgc gctggcgtcg ctgccgacca gagcatccag tacatcgagt 2460

gccacgccac cggcaccccg cgtggcgacg tcgtcgagat cgaggctgtc gagcgcgtct 2520

ttaagaagaa tgtgccccgc cttggttcta ccaagggcaa cttcggccac tccctcgtcg 2580

ctgccggctt cgcgggtatg gcgaagctgc tcctcgctat ggagcacggc gtcatcccgc 2640

ccacccctgg cctggacgcc agcaatcagg cttccgagca cgtcgtcact aaggcgatta 2700

cctggccgga aacccacggc gctcctaagc gcgccggcct ctccgccttc ggctttggcg 2760

gcacgaacgc gcatgccctt tttgaggagt tcaatgccga gggcatttcc taccgccccg 2820

gtaagcctcc ggtcgagtcg aacacccgcc cctctgtggt catcactggc atggactgca 2880

cgtttggcag cctggaggga atcgacgcct ttgagactgc cctctacgag ggccgtgatg 2940

ccgcccgtga tcttcctgcc aagcgttggc gcttcctcgg cgaggatctg gagtttctcc 3000

gcgccatccg cctgaaggag aagccccgtg gctgctttgt tgagtccgtt gacgtgaact 3060

ttcgccgcct caagacccct ctcaccccgg aggatatgct ccgcccgcag cagctcctcg 3120

ctgtctccac tatggaccgc gcgatcatcg acgccggcct caagaagggc cagcatgtcg 3180

ccgtcctcgt cggcctcggc accgatctgg agctttaccg ccaccgtgcc cgtgtggccc 3240

tcaaggaggt gctccacccc tcgctgaaga gcgacaccgc catcctccag aagatcatgc 3300

agtacgtcaa cgacgcgggc acctcgacct cctacacctc gtacatcggc aacctcgtcg 3360

ccacccgtat cagctcgcag tggggcttta ccggcccgag ctttactgtc acggagggta 3420
acaacagcgt ctaccgttgc gcccagctcg ctaaggacat gctgcaagtc aaccgcgtcg 3480
atgccgtcgt cattgccggt gtggacctga acggctccgc cgagagcttc ttcgtccgcg 3540
ctaaccgtca aaagatcagc aagctttccc acccgtgcgc cagcttcgac cgcgacgccg 3600
acggcttctt tgccggcgaa ggctgcggcg ctctcgtgtt caagcgcctt gaggactgcg 3660
cccctcagga gaagatttac gcttccatcg acagcatcgc catcgacaag gagcctacct 3720
cctcggccgt caaggctgtc taccagtccg actcctccct ctccgacatc gagcttcttg 3780
agatcagcgg cgactccaag cgttttgcgg ccttcgaggg cgctgtcgag atccagtcct 3840
ctgtggaggc gcagctcaag ggcctttcca aggttcttga gcccgccaag ggccaaggcg 3900
tcgccgttgg ctcgacccgc gccactgtcg gtgatatcgg ctacgccact ggcgcggctt 3960
cgctcatcaa gacggccctg tgcctctaca accgttacct gcctgccctc gccaactgga 4020
gcggcccctg cgaacagagc gcctggggta gcaacatgtt cgtctgccac gagactcgtc 4080
cttggatgaa gaaccagaac gagaagcgct gcgcgctcat ctcgggcacg gacccgtccc 4140
acacctgctt cagcctcgtg ctctccgata ctggttgcta cgaggagcac aaccgcacgt 4200
gctttgacgt gcaggccccg cagctcgttc tcattcatgg ttttgatggc aagaccattg 4260
tgcgccgcct tgagggctat ctccttgagc ttgtcgaggg ccatgcctcg ccgtccgagt 4320
actttcacaa gctcatcggc cagagcctcc ttgagaactc caaggagtcg aagctcacgc 4380
tctcgctcgt ctgcaacccc aaccaactcc agaaggagct catgctcgcc atcaagggtg 4440
tgcagcgctc tatgctgacg ggtaaggact gggtttcccc ctcgggctct tgctttgctc 4500
cgaaccctct ctcgtccgcc aaggtcgctt tcatgtacgg cgaaggccgc tccccctact 4560
gcggtgtggg ccttggcctc caccgcctct ggcctggcct ccacgaaaac gtgaacaaca 4620
agacggtgga cctctggacg gagggcgacg gttggctcta cccgcgcacc ctcactcgtg 4680
aggagcacac taaggccatc gagtcgttta acgctaacca gatcgagatg ttccgtgccg 4740
gtatctttat ctccatgtgc cagacggact acgtgatgaa cgtcctcggc gtgcagccca 4800
aggccggctt tggcctctcg ctcggcgaga tcagcatgct cttcgccatg tccaaggaga 4860
actgccgcca gagccaggag atgaccaacc gccttcgtgg cagccccgtc tggtcgaacg 4920
agcttgccat taacttcaat gcgatccgta agctctggaa gattccccgc ggtgccccgc 4980
ttgagtcctt ctggcagggc tacctcgtcc acggcacccg cgaggaggtc gagcacgcca 5040
tcggcctcag cgagccctac gtccgcctgc ttatcgtgaa tgacagccgc tcggccctta 5100
tcgcgggtaa gcctgacgcc tgccaggccg tgatctcgcg cctcaactcg aagttccctt 5160
ctctccctgt gaagcagggc atgatcggcc actgcccgga ggtgcgcgcc ttcatcaagg 5220
acattggtta catccacgag actctccgta tctctaacga ttactcggac tgccagctct 5280
tttcggccgt caccaagggt gcccttgact cgtctactat ggagatcaag cacttcgtcg 5340
gtgaggtcta cagccgcatt gccgattttc cccagattgt caacacggtg cactcggccg 5400

gctacgacgt gtttctcgaa ctcggctgcg acgcctcccg cagcgccgct gttcagaaca 5460

tcctcggcgg tcagggcaag tttctgtcta ctgccattga caagaagggt cactctgcct 5520

ggtcgcaggt tctccgcgct accgcctcgc tcgctgccca ccgcgtgccc ggcatctcga 5580

tcctcgatct ctttcacccg aatttccgcg agatgtgctg cacgatggct acgaccccca 5640

aggtggagga caagtttctc cgcaccatcc agattaacgg ccgtttcgag aaggagatga 5700

tccacctcga agataccact ctctcctgcc ttcccgcgcc ttcggaggcc aacattgccg 5760

ccattcagtc gcgctcgatt cgcagcgctg cggcccgtag cggccagtcg cacgactgcg 5820

cgtcccacag ccatgaggag aataaggact cctgcccgga gaagctcaaa cttgactcgg 5880

tcagcgtcgc gatcaacttt gacaacgacg accgtattca gctcggccac gcgggttttc 5940

gcgagatgta caacacccgc tactcgctct acacgggcgc tatggctaaa ggcatcgctt 6000

cggccgacct cgtcatcgct gcgggcaagg agggtatcct cgccagctac ggtgctggcg 6060

gcctcccgct cgccaccgtc cgcaagggta tcgacaagat ccagcaggcg ctccccagcg 6120

gtccttacgc cgtcaacctc atccactccc cgtttgacgg caacctggag cagggcaacg 6180

tcgacctctt tctcgaaaag aacgtccgcg tcgccgagtg ttctgccttt acgaccctca 6240

ccgtgcccgt cgtccattat cgcgccgctg gcctcgttcg ccgccaggac ggctcgatcc 6300

tcattaagaa ccgcatcatc gccaaggtca gccgcaccga gctcgctgag atgttcctcc 6360

gccctgcgcc gcagattatc ctcgaaaagc tggtcgccgc cgagatcatc tccagcgatc 6420

aggcgcgtat ggccgcgaag gtgccgatgg ccgacgacat cgccgtcgag gccgacagcg 6480

gcggccacac ggacaaccgc ccgatgcatg ttattctccc cctcattatc cagctccgca 6540

acaccattct cgcggagtac ggctgcgcta ctgcctttcg cacgcgtatt ggcgctggtg 6600

gcggcatcgg ttgcccgagc gccgctctcg ccgccttcga catgggtgcc tctttcgtcg 6660

tcaccggctc catcaatcag atttgccgcg aggcgggcac ctgcgacacg gttcgcgagc 6720

tcctcgccaa ctcgtcgtac agcgatgtca cgatggctcc tgccgccgac atgttcgatc 6780

agggcgtgaa gctccaagtc ctcaagcgcg gcaccatgtt cccctcgcgc gctaacaagc 6840

ttcgcaagct ctttgtgaac tacgagtccc tcgaaactct cccctccaag gagctcaagt 6900

accttgagaa cattattttc aagcaggctg tggaccaggt gtgggaggag actaagcgct 6960

tctactgcga gaagctgaac aaccccgaca agatcgctcg cgctatgaag gaccccaagc 7020

tcaagatgtc gctctgcttc cgctggtatc tcagcaagtc gtcgggctgg gccaacgccg 7080

gcatcaagtc tcgcgccctc gactaccaaa tctggtgcgg acccgcgatg ggcagcttta 7140

acaacttcgc ctccggcacg agcctcgact ggaaagtgac cggtgtcttt cccggcgtcg 7200

ccgaggtgaa catggccatc ctcgatggcg cccgcgagct cgctgccaag cgcaactaag 7260

taagctctca tatatgttag ttgcgtgaga ccgacacgaa gataatatca catacgcttt 7320

tgtttgttct ttcaattatt tgtctgtgct tcatgttgct cctcagtatc tagctggcgg 7380

ctcttatctt cttttaaaat atctggacaa ggacaaaaac aagaataaag gcgagaagat 7440

gtgaatttca tttcgacttg agaactcgaa gagcattgat gcggttagta tatgggtatt 7500

ttccagacac ttttcatcat catcatcatc atcatcatta tgaagaagta gtagctgata 7560

aagtagactc actgtttgca gcgagaacca ggccctttcg gtccgtaacc ttggtgaaaa 7620

atcgaggagc gatcatgcct tgtgcgtcgc cggggctaag aatgaccccg ttgtctgatt 7680

cgaaaaattt gatgccagcc tcggcagctg cagcagcagc gtctacatac acgaagtacg 7740

tcacaaggac ttgcgcatcc cgctgatcac acccgtctcc gaggacggaa gcccgtgagc 7800

aaagtgaatg tgattccgtt tcattcgcga caaacccttg tcctggatct gagcccaggc 7860

agaagtattt gttccgtgca agcacgtctc aatgcttgct ggatcgagtg gacgcaacaa 7920

actcgtctct ttgacaactg acatggagtg tccctgattc gcacgaatca accagcttcc 7980

gtccgtgtcc tggcgcaacg tgaaccgctg cttgtcgttg tctcgcacga cagcctcaat 8040

gtccgaaaag gagatcccct tgaagcgctt gagcttgccc tgtagctcag ccagacgaac 8100

aaatccatct tcgcgcactt cgagaccctc ctgaacggcc gagtgccgca gcacacggct 8160

tagcatgtgc gaaagcttct tgttatgcgc cttggctttt gaaggcagcg ttggcgcgct 8220

gcgagagtgt ttacttggtt gcgcagcttc actatcacga gcagcacttc gtttattcat 8280

ttccatgttg tttgcaggtt cagtggtagc cccatccggc acgggccaga gcaccctgtg 8340

agagcgccgt tcctcgaccc tgaggaaaag ctccgtaggt atggctcctg tatcgcccag 8400

accaggggta agtaggacgc cattgcttga ttggaacatt tccaagcccg catcatgtgc 8460

gcgcgcggta tcgaggaaga tgagcacctc tgcggagccg cgaaagccac tgagcactcc 8520

gtcctgacca ggaagtcccg aggcaaaatg aatgtgattt cggcccatgc ggcggaggcc 8580

ctggcgttgg attgcaggcc agtggcgaag atacgttccg tgaatacaca gcggaagcaa 8640

tgagaga 8647

<210> 73
<211> 4479
<212> DNA
<213> Artificial

<220>
<223> synthetic

<220>
<221> CDS
<222> (1)..(4479)

<400> 73
atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc 48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1 5 10 15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa 96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
20 25 30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
35 40 45

```
ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc      192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac      240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc      288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc      336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac      384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc      432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att      480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc      528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg      576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc      624
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc      672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag      720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg      768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
                245                 250                 255

gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa      816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
            260                 265                 270

ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac      864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
        275                 280                 285

cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc      912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
    290                 295                 300

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag      960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305                 310                 315                 320
```

```
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag     1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
                325                 330                 335

atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac     1056
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
            340                 345                 350

ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa     1104
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365

atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag     1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
    370                 375                 380

atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac     1200
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385                 390                 395                 400

atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc     1248
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
                405                 410                 415

atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac     1296
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
            420                 425                 430

ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac     1344
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
        435                 440                 445

ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc     1392
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
    450                 455                 460

ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc     1440
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
465                 470                 475                 480

gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc     1488
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
                485                 490                 495

gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag     1536
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
            500                 505                 510

gag atg tcc tgg cac ccc atg tca aag ctt cct ggc aac cct acg ccg     1584
Glu Met Ser Trp His Pro Met Ser Lys Leu Pro Gly Asn Pro Thr Pro
        515                 520                 525

ttc ttc tcc cct tca tct tac cct ccg agg gca att tgc ttc atc cct     1632
Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys Phe Ile Pro
    530                 535                 540

ttc ccg ggc aat ccc ctt gac aac aac tgc aag gct gga gaa atg ccc     1680
Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly Glu Met Pro
545                 550                 555                 560

ctg aac tgg tac aac atg tca gag ttc atg tgt ggc aag gtt tct aac     1728
Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys Val Ser Asn
                565                 570                 575

tgc ttg ggc cca gaa ttc gca cgc ttt gac aag tcg aac acc agc cgg     1776
Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn Thr Ser Arg
            580                 585                 590
```

```
agc cct gct ttt gac ttg gct ctg gtg acc cga gtt gtt gaa gtc aca     1824
Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val Glu Val Thr
        595                 600                 605

aac atg gaa cac ggc aag ttt cta aac gtt gat tgc aat cca agc aaa     1872
Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn Pro Ser Lys
    610                 615                 620

ggc aca atg gtg ggg gag ttt gac tgt ccc caa gac gcg tgg ttc ttt     1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala Trp Phe Phe
625                 630                 635                 640

gat ggt tcg tgc aac gac ggc cat atg ccg tat tcc att atc atg gaa     1968
Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile Ile Met Glu
                645                 650                 655

atc gga ctg caa acc tca ggt gtt ctc acc tcg gtg ttg aag gca ccg     2016
Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
            660                 665                 670

ctg act atg gac aag gat gac att ctc ttt cga aac ctc gat gca agt     2064
Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Ser
        675                 680                 685

gct gaa atg gtg cgt cca gac gtg gat gtt cgc ggc aaa acg att cga     2112
Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys Thr Ile Arg
    690                 695                 700

aac gtg acc aag tgt acc ggc tat gca atg ttg gga aag atg ggg att     2160
Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys Met Gly Ile
705                 710                 715                 720

cac cgg ttc acg ttt gag ttg agc gtt gac ggc gtg gta ttt tat aaa     2208
His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val Phe Tyr Lys
                725                 730                 735

gga tcc act tcc ttt gga tgg ttc act ccc gag gtg ttt gct cag caa     2256
Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe Ala Gln Gln
            740                 745                 750

gct gga ctc gac aac ggg aaa aag acg gag ccc tgg tgc aag act aac     2304
Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys Lys Thr Asn
        755                 760                 765

aac acc tcg gtt cga aga gtt gaa atc gca tcc gcc aaa gga aaa gag     2352
Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys Gly Lys Glu
    770                 775                 780

cag ctg act gag aag ctt ccc gac gca act aat gct caa gtt ctt cgg     2400
Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln Val Leu Arg
785                 790                 795                 800

cgt tca gag cag tgt gaa tac ctc gat tac ctc aat att gcc cct gac     2448
Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile Ala Pro Asp
                805                 810                 815

tct ggg ctg cat ggg aag ggc tac gcc cac gga cac aaa gac gtt aac     2496
Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys Asp Val Asn
            820                 825                 830

ccg caa gac tgg ttc ttc tct tgc cac ttt tgg ttc gat cct gta atg     2544
Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Pro Val Met
        835                 840                 845

cca gga tct tta gga att gaa tca atg ttc cag ctt atc gag gcc ttt     2592
Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile Glu Ala Phe
    850                 855                 860
```

```
gcg gtg gac caa aac att cct gga gag tac aac gta tcc aat ccg acc     2640
Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser Asn Pro Thr
865                 870                 875                 880

ttt gcc cat gca cca ggc aaa acg gcg tgg aaa tac cga ggc cag ctc     2688
Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg Gly Gln Leu
                885                 890                 895

aca cca aag aac cgt gcg atg gac tgc gag gtg cat atc gtt tca att     2736
Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile Val Ser Ile
            900                 905                 910

acc gcc tcc ccc gag aac ggg ggc tac gtt gac atc gtg gcc gat gga     2784
Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val Ala Asp Gly
        915                 920                 925

gcg ctt tgg gta gat gga ctt cgc gtg tac gaa gcc aaa gag ctt cga     2832
Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys Glu Leu Arg
    930                 935                 940

gtt cgt gtc gtt tcg gca aaa cct caa gca att ccg gat gta caa caa     2880
Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp Val Gln Gln
945                 950                 955                 960

cag cca cct agc gca aag gcg gac ccg ggg aaa aca gga gtt gca ctt     2928
Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly Val Ala Leu
                965                 970                 975

tcg ccc act cag cta cgc gac gtc ctg ctt gaa gtg gac aat cca ttg     2976
Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp Asn Pro Leu
            980                 985                 990

tat ctt ggt gta gag aac tcc aat  ttg gtg cag ttt gag  tcg aaa cct   3024
Tyr Leu Gly Val Glu Asn Ser Asn  Leu Val Gln Phe Glu  Ser Lys Pro
        995                 1000                1005

gca act  tct tca cgt atc gtt  tcg atc aaa ccg tgc  tcg att agt      3069
Ala Thr  Ser Ser Arg Ile Val  Ser Ile Lys Pro Cys  Ser Ile Ser
    1010                1015                1020

gac ctt  ggc gat aag tct ttt  atg gaa acg tac aac  gtg tca gca      3114
Asp Leu  Gly Asp Lys Ser Phe  Met Glu Thr Tyr Asn  Val Ser Ala
    1025                1030                1035

cct ctg  tat act ggc gcc atg  gcc aag ggc att gcc  tcg gcg gac      3159
Pro Leu  Tyr Thr Gly Ala Met  Ala Lys Gly Ile Ala  Ser Ala Asp
    1040                1045                1050

ctc gtc  atc gcc gcc ggc aag  cgc aag atc ctc ggc  tcc ttt ggc      3204
Leu Val  Ile Ala Ala Gly Lys  Arg Lys Ile Leu Gly  Ser Phe Gly
    1055                1060                1065

gcc ggc  ggc ctc ccc atg cac  cac gtg cgc gcc gcc  ctc gag aag      3249
Ala Gly  Gly Leu Pro Met His  His Val Arg Ala Ala  Leu Glu Lys
    1070                1075                1080

atc cag  gcc gcc ctg cct cag  ggc ccc tac gcc gtc  aac ctc atc      3294
Ile Gln  Ala Ala Leu Pro Gln  Gly Pro Tyr Ala Val  Asn Leu Ile
    1085                1090                1095

cac tcg  cct ttt gac agc aac  ctc gag aag ggc aac  gtc gat ctc      3339
His Ser  Pro Phe Asp Ser Asn  Leu Glu Lys Gly Asn  Val Asp Leu
    1100                1105                1110

ttc ctc  gag aag ggc gtc act  gtg gtg gag gcc tcg  gca ttc atg      3384
Phe Leu  Glu Lys Gly Val Thr  Val Val Glu Ala Ser  Ala Phe Met
    1115                1120                1125
```

```
acc ctc  acc ccg cag gtc gtg  cgc tac cgc gcc gcc  ggc ctc tcg      3429
Thr Leu  Thr Pro Gln Val Val  Arg Tyr Arg Ala Ala  Gly Leu Ser
    1130                 1135                 1140

cgc aac  gcc gac ggt tcg gtc  aac atc cgc aac cgc  atc atc ggc      3474
Arg Asn  Ala Asp Gly Ser Val  Asn Ile Arg Asn Arg  Ile Ile Gly
    1145                 1150                 1155

aag gtc  tcg cgc acc gag ctc  gcc gag atg ttc atc  cgc ccg gcc      3519
Lys Val  Ser Arg Thr Glu Leu  Ala Glu Met Phe Ile  Arg Pro Ala
    1160                 1165                 1170

ccg gag  cac ctc ctc gag aag  ctc atc gcc tcg ggc  gag atc acc      3564
Pro Glu  His Leu Leu Glu Lys  Leu Ile Ala Ser Gly  Glu Ile Thr
    1175                 1180                 1185

cag gag  cag gcc gag ctc gcg  cgc cgc gtt ccc gtc  gcc gac gat      3609
Gln Glu  Gln Ala Glu Leu Ala  Arg Arg Val Pro Val  Ala Asp Asp
    1190                 1195                 1200

atc gct  gtc gag gct gac tcg  ggc ggc cac acc gac  aac cgc ccc      3654
Ile Ala  Val Glu Ala Asp Ser  Gly Gly His Thr Asp  Asn Arg Pro
    1205                 1210                 1215

atc cac  gtc atc ctc ccg ctc  atc atc aac ctc cgc  aac cgc ctg      3699
Ile His  Val Ile Leu Pro Leu  Ile Ile Asn Leu Arg  Asn Arg Leu
    1220                 1225                 1230

cac cgc  gag tgc ggc tac ccc  gcg cac ctc cgc gtc  cgc gtt ggc      3744
His Arg  Glu Cys Gly Tyr Pro  Ala His Leu Arg Val  Arg Val Gly
    1235                 1240                 1245

gcc gga  ggt ggc gtc ggc tgc  ccg cag gcc gcc gcc  gcc gcg ctc      3789
Ala Gly  Gly Gly Val Gly Cys  Pro Gln Ala Ala Ala  Ala Ala Leu
    1250                 1255                 1260

acc atg  ggc gcc gcc ttc atc  gtc acc ggc act gtc  aac cag gtc      3834
Thr Met  Gly Ala Ala Phe Ile  Val Thr Gly Thr Val  Asn Gln Val
    1265                 1270                 1275

gcc aag  cag tcc ggc acc tgc  gac aac gtg cgc aag  cag ctc tcg      3879
Ala Lys  Gln Ser Gly Thr Cys  Asp Asn Val Arg Lys  Gln Leu Ser
    1280                 1285                 1290

cag gcc  acc tac tcg gat atc  tgc atg gcc ccg gcc  gcc gac atg      3924
Gln Ala  Thr Tyr Ser Asp Ile  Cys Met Ala Pro Ala  Ala Asp Met
    1295                 1300                 1305

ttc gag  gag ggc gtc aag ctc  cag gtc ctc aag aag  gga acc atg      3969
Phe Glu  Glu Gly Val Lys Leu  Gln Val Leu Lys Lys  Gly Thr Met
    1310                 1315                 1320

ttc ccc  tcg cgc gcc aac aag  ctc tac gag ctc ttt  tgc aag tac      4014
Phe Pro  Ser Arg Ala Asn Lys  Leu Tyr Glu Leu Phe  Cys Lys Tyr
    1325                 1330                 1335

gac tcc  ttc gac tcc atg cct  cct gcc gag ctc gag  cgc atc gag      4059
Asp Ser  Phe Asp Ser Met Pro  Pro Ala Glu Leu Glu  Arg Ile Glu
    1340                 1345                 1350

aag cgt  atc ttc aag cgc gca  ctc cag gag gtc tgg  gag gag acc      4104
Lys Arg  Ile Phe Lys Arg Ala  Leu Gln Glu Val Trp  Glu Glu Thr
    1355                 1360                 1365

aag gac  ttt tac att aac ggt  ctc aag aac ccg gag  aag atc cag      4149
Lys Asp  Phe Tyr Ile Asn Gly  Leu Lys Asn Pro Glu  Lys Ile Gln
    1370                 1375                 1380
```

```
cgc gcc  gag cac gac ccc aag  ctc aag atg tcg ctc  tgc ttc cgc      4194
Arg Ala  Glu His Asp Pro Lys  Leu Lys Met Ser Leu  Cys Phe Arg
    1385                 1390                 1395

tgg tac  ctt ggt ctt gcc agc  cgc tgg gcc aac atg  ggc gcc ccg      4239
Trp Tyr  Leu Gly Leu Ala Ser  Arg Trp Ala Asn Met  Gly Ala Pro
    1400                 1405                 1410

gac cgc  gtc atg gac tac cag  gtc tgg tgt ggc ccg  gcc att ggc      4284
Asp Arg  Val Met Asp Tyr Gln  Val Trp Cys Gly Pro  Ala Ile Gly
    1415                 1420                 1425

gcc ttc  aac gac ttc atc aag  ggc acc tac ctc gac  ccc gct gtc      4329
Ala Phe  Asn Asp Phe Ile Lys  Gly Thr Tyr Leu Asp  Pro Ala Val
    1430                 1435                 1440

tcc aac  gag tac ccc tgt gtc  gtc cag atc aac ctg  caa atc ctc      4374
Ser Asn  Glu Tyr Pro Cys Val  Val Gln Ile Asn Leu  Gln Ile Leu
    1445                 1450                 1455

cgt ggt  gcc tgc tac ctg cgc  cgt ctc aac gcc ctg  cgc aac gac      4419
Arg Gly  Ala Cys Tyr Leu Arg  Arg Leu Asn Ala Leu  Arg Asn Asp
    1460                 1465                 1470

ccg cgc  att gac ctc gag acc  gag gat gct gcc ttt  gtc tac gag      4464
Pro Arg  Ile Asp Leu Glu Thr  Glu Asp Ala Ala Phe  Val Tyr Glu
    1475                 1480                 1485

ccc acc  aac gcg ctc                                             4479
Pro Thr  Asn Ala Leu
    1490
```

<210> 74
<211> 1493
<212> PRT
<213> Artificial

<220>
<223> Synthetic Construct

<400> 74

```
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1               5                  10                  15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
            20                  25                  30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
        35                  40                  45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
    50                  55                  60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65                  70                  75                  80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
                85                  90                  95
```

```
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
            100                 105                 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
        115                 120                 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
    130                 135                 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145                 150                 155                 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
                165                 170                 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
            180                 185                 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
        195                 200                 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
    210                 215                 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225                 230                 235                 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
                245                 250                 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
            260                 265                 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
        275                 280                 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
    290                 295                 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305                 310                 315                 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
                325                 330                 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
            340                 345                 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
        355                 360                 365
```

```
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
    370                 375                 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385                 390                 395                 400

Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
                405                 410                 415

Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
            420                 425                 430

Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
        435                 440                 445

Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
    450                 455                 460

Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
465                 470                 475                 480

Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
                485                 490                 495

Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
            500                 505                 510

Glu Met Ser Trp His Pro Met Ser Lys Leu Pro Gly Asn Pro Thr Pro
        515                 520                 525

Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys Phe Ile Pro
    530                 535                 540

Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly Glu Met Pro
545                 550                 555                 560

Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys Val Ser Asn
                565                 570                 575

Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn Thr Ser Arg
            580                 585                 590

Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val Glu Val Thr
        595                 600                 605

Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn Pro Ser Lys
    610                 615                 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala Trp Phe Phe
625                 630                 635                 640
```

```
Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile Ile Met Glu
                645                 650                 655

Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
            660                 665                 670

Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Ser
        675                 680                 685

Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys Thr Ile Arg
    690                 695                 700

Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys Met Gly Ile
705                 710                 715                 720

His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val Phe Tyr Lys
                725                 730                 735

Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe Ala Gln Gln
            740                 745                 750

Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys Lys Thr Asn
        755                 760                 765

Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys Gly Lys Glu
    770                 775                 780

Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln Val Leu Arg
785                 790                 795                 800

Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile Ala Pro Asp
                805                 810                 815

Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys Asp Val Asn
            820                 825                 830

Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Pro Val Met
        835                 840                 845

Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile Glu Ala Phe
    850                 855                 860

Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser Asn Pro Thr
865                 870                 875                 880

Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg Gly Gln Leu
                885                 890                 895

Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile Val Ser Ile
            900                 905                 910
```

```
Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val Ala Asp Gly
        915                 920                 925

Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys Glu Leu Arg
    930                 935                 940

Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp Val Gln Gln
945                 950                 955                 960

Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly Val Ala Leu
                965                 970                 975

Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp Asn Pro Leu
            980                 985                 990

Tyr Leu Gly Val Glu Asn Ser Asn  Leu Val Gln Phe Glu  Ser Lys Pro
        995                 1000                1005

Ala Thr  Ser Ser Arg Ile Val  Ser Ile Lys Pro Cys  Ser Ile Ser
    1010                1015                1020

Asp Leu  Gly Asp Lys Ser Phe  Met Glu Thr Tyr Asn  Val Ser Ala
    1025                1030                1035

Pro Leu  Tyr Thr Gly Ala Met  Ala Lys Gly Ile Ala  Ser Ala Asp
    1040                1045                1050

Leu Val  Ile Ala Ala Gly Lys  Arg Lys Ile Leu Gly  Ser Phe Gly
    1055                1060                1065

Ala Gly  Gly Leu Pro Met His  His Val Arg Ala Ala  Leu Glu Lys
    1070                1075                1080

Ile Gln  Ala Ala Leu Pro Gln  Gly Pro Tyr Ala Val  Asn Leu Ile
    1085                1090                1095

His Ser  Pro Phe Asp Ser Asn  Leu Glu Lys Gly Asn  Val Asp Leu
    1100                1105                1110

Phe Leu  Glu Lys Gly Val Thr  Val Val Glu Ala Ser  Ala Phe Met
    1115                1120                1125

Thr Leu  Thr Pro Gln Val Val  Arg Tyr Arg Ala Ala  Gly Leu Ser
    1130                1135                1140

Arg Asn  Ala Asp Gly Ser Val  Asn Ile Arg Asn Arg  Ile Ile Gly
    1145                1150                1155

Lys Val  Ser Arg Thr Glu Leu  Ala Glu Met Phe Ile  Arg Pro Ala
    1160                1165                1170
```

```
Pro Glu  His Leu Leu Glu Lys  Leu Ile Ala Ser Gly  Glu Ile Thr
    1175                 1180                 1185

Gln Glu  Gln Ala Glu Leu Ala  Arg Arg Val Pro Val  Ala Asp Asp
    1190                 1195                 1200

Ile Ala  Val Glu Ala Asp Ser  Gly Gly His Thr Asp  Asn Arg Pro
    1205                 1210                 1215

Ile His  Val Ile Leu Pro Leu  Ile Ile Asn Leu Arg  Asn Arg Leu
    1220                 1225                 1230

His Arg  Glu Cys Gly Tyr Pro  Ala His Leu Arg Val  Arg Val Gly
    1235                 1240                 1245

Ala Gly  Gly Gly Val Gly Cys  Pro Gln Ala Ala Ala  Ala Ala Leu
    1250                 1255                 1260

Thr Met  Gly Ala Ala Phe Ile  Val Thr Gly Thr Val  Asn Gln Val
    1265                 1270                 1275

Ala Lys  Gln Ser Gly Thr Cys  Asp Asn Val Arg Lys  Gln Leu Ser
    1280                 1285                 1290

Gln Ala  Thr Tyr Ser Asp Ile  Cys Met Ala Pro Ala  Ala Asp Met
    1295                 1300                 1305

Phe Glu  Glu Gly Val Lys Leu  Gln Val Leu Lys Lys  Gly Thr Met
    1310                 1315                 1320

Phe Pro  Ser Arg Ala Asn Lys  Leu Tyr Glu Leu Phe  Cys Lys Tyr
    1325                 1330                 1335

Asp Ser  Phe Asp Ser Met Pro  Pro Ala Glu Leu Glu  Arg Ile Glu
    1340                 1345                 1350

Lys Arg  Ile Phe Lys Arg Ala  Leu Gln Glu Val Trp  Glu Glu Thr
    1355                 1360                 1365

Lys Asp  Phe Tyr Ile Asn Gly  Leu Lys Asn Pro Glu  Lys Ile Gln
    1370                 1375                 1380

Arg Ala  Glu His Asp Pro Lys  Leu Lys Met Ser Leu  Cys Phe Arg
    1385                 1390                 1395

Trp Tyr  Leu Gly Leu Ala Ser  Arg Trp Ala Asn Met  Gly Ala Pro
    1400                 1405                 1410

Asp Arg  Val Met Asp Tyr Gln  Val Trp Cys Gly Pro  Ala Ile Gly
    1415                 1420                 1425
```

Ala Phe Asn Asp Phe Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val
1430 1435 1440

Ser Asn Glu Tyr Pro Cys Val Val Gln Ile Asn Leu Gln Ile Leu
1445 1450 1455

Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp
1460 1465 1470

Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu
1475 1480 1485

Pro Thr Asn Ala Leu
1490

<210> 75
<211> 4479
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 75
atggcgctcc gtgtcaagac gaacaagaag ccatgctggg agatgaccaa ggaggagctg 60

accagcggca agaccgaggt gttcaactat gaggaactcc tcgagttcgc agagggcgac 120

atcgccaagg tcttcggacc cgagttcgcc gtcatcgaca agtacccgcg ccgcgtgcgc 180

ctgcccgccc gcgagtacct gctcgtgacc cgcgtcaccc tcatggacgc cgaggtcaac 240

aactaccgcg tcggcgcccg catggtcacc gagtacgatc tccccgtcaa cggagagctc 300

tccgagggcg gagactgccc ctgggccgtc ctggtcgaga gtggccagtg cgatctcatg 360

ctcatctcct acatgggcat tgacttccag aaccagggcg accgcgtcta ccgcctgctc 420

aacaccacgc tcacctttta cggcgtggcc cacgagggcg agaccctcga gtacgacatt 480

cgcgtcaccg gcttcgccaa gcgtctcgac ggcggcatct ccatgttctt cttcgagtac 540

gactgctacg tcaacggccg cctcctcatc gagatgcgcg atggctgcgc cggcttcttc 600

accaacgagg agctcgacgc cggcaagggc gtcgtcttca cccgcggcga cctcgccgcc 660

cgcgccaaga tcccaaagca ggacgtctcc ccctacgccg tcgccccctg cctccacaag 720

accaagctca acgaaaagga gatgcagacc ctcgtcgaca aggactgggc atccgtcttt 780

ggctccaaga acggcatgcc ggaaatcaac tacaaactct gcgcgcgtaa gatgctcatg 840

attgaccgcg tcaccagcat tgaccacaag ggcggtgtct acggcctcgg tcagctcgtc 900

ggtgaaaaga tcctcgagcg cgaccactgg tactttccct gccactttgt caaggatcag 960

gtcatggccg gatccctcgt ctccgacggc tgcagccaga tgctcaagat gtacatgatc 1020

tggctcggcc tccacctcac caccggaccc tttgacttcc gcccggtcaa cggccacccc 1080

aacaaggtcc gctgccgcgg ccaaatctcc ccgcacaagg gcaagctcgt ctacgtcatg 1140

gagatcaagg agatgggctt cgacgaggac aacgacccgt acgccattgc cgacgtcaac 1200

atcattgatg tcgacttcga aaagggccag gactttagcc tcgaccgcat cagcgactac 1260

ggcaagggcg acctcaacaa gaagatcgtc gtcgacttta agggcatcgc tctcaagatg 1320

cagaagcgct ccaccaacaa gaacccctcc aaggttcagc ccgtctttgc caacggcgcc 1380

gccactgtcg gccccgaggc ctccaaggct tcctccggcg ccagcgccag cgccagcgcc 1440

gccccggcca agcctgcctt cagcgccgat gttcttgcgc ccaagcccgt tgcccttccc 1500

gagcacatcc tcaagggcga cgccctcgcc cccaaggaga tgtcctggca ccccatgagc 1560

aagctccccg gcaaccccac ccccttcttt tctccctcgt cgtatccgcc ccgtgctatc 1620

tgcttcatcc cttttccggg caacccgctc gacaacaact gcaaggccgg cgagatgccc 1680

ctcaactggt acaacatgag cgaattcatg tgcggtaagg tcagcaactg cctcggcccg 1740

gagttcgccc gctttgacaa gagcaacacc tcccgctccc ccgccttcga ccttgcgctc 1800

gtcacgcgcg tcgtcgaggt caccaacatg gagcacggca agttcctcaa cgtcgactgt 1860

aacccttcga agggcacgat ggtgggagag ttcgactgcc cgcaggacgc ctggttcttt 1920

gacggctcct gcaatgacgg tcatatgccg tactccatca tcatggagat cggccttcag 1980

acgtctggtg tcctcacgtc ggtgctcaag gctcctctca cgatggacaa ggacgacatc 2040

ctctttcgca accttgatgc ctcggctgag atggtgcgcc ctgacgtcga tgtccgcggc 2100

aagaccatcc gcaacgtgac caagtgtacc ggctacgcca tgcttggtaa gatgggtatc 2160

catcgcttta cctttgagct ctccgtggac ggcgtcgtct tttacaaggg cagcaccagc 2220

ttcggctggt ttacgccgga agtgtttgcc cagcaggccg gtctggacaa cggcaagaag 2280

accgagccct ggtgcaagac caacaacacc tccgtccgcc gcgtcgagat cgccagcgct 2340

aagggtaagg agcagctcac ggagaagctc cctgatgcta ccaacgccca ggttctccgc 2400

cgctccgagc agtgcgagta cctcgactac ctcaacatcg cccccgactc cggccttcac 2460

ggcaagggtt acgcgcacgg ccacaaggac gtcaaccccc aggactggtt tttctcgtgc 2520

cacttctggt tcgacccggt gatgcccggc tcgcttggca ttgagtccat gttccagctc 2580

atcgaagcct ttgctgtcga ccagaatatc cctggcgagt acaatgtttc gaacccgacc 2640

tttgcccatg cccccggtaa gaccgcttgg aagtatcgcg gccagctcac tcccaagaac 2700

cgcgctatgg actgcgaggt ccacatcgtc agcatcacgg cgagcccgga gaacggcggc 2760

tacgtcgaca tcgtcgccga cggcgccctt tgggtcgatg gcctccgcgt ttacgaggcc 2820

aaggagctcc gtgtccgcgt cgtcagcgcc aagccgcaag ctatcccgga tgtccagcag 2880

cagcctccct cggctaaggc cgaccctggc aagaccggtg tcgccctcag ccccacccag 2940

ctccgcgatg tcctcctgga ggtcgacaac cctctgtacc tcggtgtcga gaactccaac 3000

cttgtccagt tcgaaagcaa gccggcgacg tcctcccgca tcgtgtccat taagccgtgt 3060

tctatctccg acctcggcga caagtctttc atggagactt acaacgtgtc tgcgcctctg 3120

tatactggcg ccatggccaa gggcattgcc tcggcggacc tcgtcatcgc cgccggcaag 3180

cgcaagatcc tcggctcctt tggcgccggc ggcctcccca tgcaccacgt gcgcgccgcc 3240

ctcgagaaga tccaggccgc cctgcctcag ggcccctacg ccgtcaacct catccactcg 3300

ccttttgaca gcaacctcga gaagggcaac gtcgatctct tcctcgagaa gggcgtcact 3360

gtggtggagg cctcggcatt catgaccctc accccgcagg tcgtgcgcta ccgcgccgcc 3420

ggcctctcgc gcaacgccga cggttcggtc aacatccgca accgcatcat cggcaaggtc 3480

tcgcgcaccg agctcgccga gatgttcatc cgcccggccc cggagcacct cctcgagaag 3540

ctcatcgcct cgggcgagat cacccaggag caggccgagc tcgcgcgccg cgttcccgtc 3600

gccgacgata tcgctgtcga ggctgactcg ggcggccaca ccgacaaccg ccccatccac 3660

gtcatcctcc cgctcatcat caacctccgc aaccgcctgc accgcgagtg cggctacccc 3720

gcgcacctcc gcgtccgcgt tggcgccgga ggtggcgtcg gctgcccgca ggccgccgcc 3780

gccgcgctca ccatgggcgc cgccttcatc gtcaccggca ctgtcaacca ggtcgccaag 3840

cagtccggca cctgcgacaa cgtgcgcaag cagctctcgc aggccaccta ctcggatatc 3900

tgcatggccc cggccgccga catgttcgag gagggcgtca agctccaggt cctcaagaag 3960

ggaaccatgt tcccctcgcg cgccaacaag ctctacgagc tcttttgcaa gtacgactcc 4020

ttcgactcca tgcctcctgc cgagctcgag cgcatcgaga agcgtatctt caagcgcgca 4080

ctccaggagg tctgggagga gaccaaggac ttttacatta acggtctcaa gaacccggag 4140

aagatccagc gcgccgagca cgaccccaag ctcaagatgt cgctctgctt ccgctggtac 4200

cttggtcttg ccagccgctg ggccaacatg ggcgccccgg accgcgtcat ggactaccag 4260

gtctggtgtg gcccggccat tggcgccttc aacgacttca tcaagggcac ctacctcgac 4320

cccgctgtct ccaacgagta cccctgtgtc gtccagatca acctgcaaat cctccgtggt 4380

gcctgctacc tgcgccgtct caacgccctg cgcaacgacc cgcgcattga cctcgagacc 4440

gaggatgctg cctttgtcta cgagcccacc aacgcgctc 4479

<210> 76
<211> 4200
<212> DNA
<213> Schizochytrium sp.

<400> 76
tttctctctc tcgagctgtt gctgctgctg ctgctgctgc tgcttccttg ctggttctca 60

cgtccgttcg atcaagcgct cgctcgctcg accgatcggt gcgtgcgtgc gtgcgtgagt 120

cttgttgcca ggcagccgca ggctgtctgt ctgtttgtgt agttttaccc tcggggttcg 180

gggtctgcct gcctcccgct cccgcccgcc gccgcccgta tccaccccgc tcgcctccgc 240

ccatcgggcc tcgcctcctc gcgccgcacg catcgcgcgc atcgcatgca tcatgctgcc 300

acgcacgggg ggacgcgcgc cccgcgtccc ccgccgccgc cgtcgtcgtc tggcgatgcc 360

gtcgccgccc tccttccttc cctcgcctcc tcttcctccc gagcccccct gtcttccttc 420

gcccccgcag cggcgcgcag gaagcgagga gagcggggag gagagaagaa aagaaaagaa 480

aagaaaagaa aataacagcg ccgtctcgcg cagacgcgcg cggccgcgtg cgaggcggcg 540

tgatggggct tctcgtggcg cggctgcggc ctggcccggc ctcgcctttg aggtgcaggc 600

tttgggagag aagagtggga cgcggagaag ataagatggt gccatggcgc aggacggaga 660

ggttgctgaa acttcttcga gcggcacagg cgatggcgag agaccgacag ctgccggcgc 720

ggaggggatg gatacctccc gaggctggca tggacgagct ggccgcgcgg atctggctgg 780

ccgcgcggcg gtgggtccgg aggcgcgagg ttggttttct tcatacctga taccatacgg 840

tattcattct tcctctccag gaaggaagca agtcacatag agtatcacta gcctaatgat 900

ggactctatg ttttagggca cgtcggagca gaaggcgcga gcgattcgaa tgcgagcgat 960

agatacagca cagagacctt gccggcgacg cggatgcagg cgagcacgca cgcaccgcac 1020

gcacggcagc ggtgcacgcg ctcctcggca gatgcacggt tctgcgccgc gcctttacat 1080

tttttgattt taggtggtgt gcctgccact ttgaacatca tccacaagtc aacgcagcat 1140

caagaggcaa gcaagtacat acatccattc gaattcaagt tcaagagacg cagcaacagc 1200

cgccgctccg ctcaagctgc agctagctgg ctgacagggc tcgctggctg tagtggaaaa 1260

ttccattcac ttttctgcat ccgcggccag caggcccgta cgcacgttct ctcgtttgtt 1320

tgttcgttcg tgcgtgcgtg cgtgcgtccc agctgcctgt ctaatctgcc gcgcgatcca 1380

acgaccctcg gtcgtcgccg caagcgaaac ccgacgccga cctggccaat gccgcaagaa 1440

tgctaagcgc gcagcaatgc tgagagtaat cttcagccca ccaagtcatt atcgctgccc 1500

aagtctccat cgcagccaca ttcaggcttt ctctctctct ccctccctct ctttctgccg 1560

ggagagaagg aaagacccgc cgccgccgcc tctgcgcctg tgacgggctg tccgttgtaa 1620

gccctcttag acagttccta ggtgccgggc gccgccgcgc ctccgtcgca ggcacacgta 1680

ggcggccacg ggttcccccc gcaccttcca caccttcttc ccccgcagcc ggaccgcgcg 1740

ccgtctgctt acgcacttcg cgcggccgcc gcccgcgaac ccgagcgcgt gctgtgggcg 1800

ccgtcttccg gccgcgtcgg aggtcgtccc cgcgccgcgc tactccgggt cctgtgcggt 1860

acgtacttaa tattaacagt gggacctcgc acaggacctg acggcagcac agacgtcgcc 1920

gcctcgcatc gctggggacg caggcgaggc atcccggcgc ggccccgcac cggggaggct 1980

gcggggcggc ctcttccggc cggcggccgc atcaggcgga tgacgcaaga gccctcgcag 2040

tcgctcgctc gcgggagcgc agcgcggcgc cagcgtggcc aagctcccgc cccttctggc 2100

tggctgcatg cctgcctgcc tgcctgcctg cgtgcgtgcg tgcgtgcgtg ccttcgtgcg 2160

tgcctgcctt cgtgcgtgcg tgcgtgagtg cggcggaaga gggatcatgc gaggatcaat 2220

cacccgccgc acctcgactt ttgaagaagc cgcgatgcga tgcgatgcga tgcgatgcga 2280

cgcgataccg tgcgaggcta cgaagcgagt ctggccggcc gtcatacaac gcacgttttc 2340

gagaaggagg gctggcggag gcgtgcatgc cggcgaccat tgcgaacgcg gcgtctcgtg 2400

gctggcgaag gtgcctggag gatctaacga tcgctgctat gatgctatag ctgtgctgat 2460

ccccggtcca ttccaccacg tctgtgcctg ccgcctgacc tgcgcttggc tttccttcaa 2520

gttctcctcc gccgggcctt caggaccgag acgagacctg cagctgcagc tagactcgcg 2580

ctcgctcgcg gaggattcgc cggccgccgg gccggacggg actcgcgagg tcacacggcc 2640

gccggcgatc gcgatggctg tgctgacgta ctcgtgcgtg gcagccgtac gtcagcgacg 2700

ccgcctccgt attgtggatt cgttagttgg ttgttggttg atttgttgat taatttttt 2760

gttcgtaggc ttggttatag ctaatagttt agtttatact ggtgctcttc ggtgctgatt 2820

tagctcgact tgggtccaca ccactgcccc tctactgtga atggatcaat ggacgcacga 2880

cgggccgacg aaagtgcgcg agtgaggtaa cctaagcaac ggcggtcttc agaggggacg 2940

cacgccctcc gtcgcagtca gtccagacag gcagaaaagc gtcttaggga ccacgcacgc 3000

acgcacgcac gcacgcacgc ccgcacgcac gctccctccc tcgcgtgcct atttttttag 3060

gcttccttcc gcacgggcct acctctcgct ccctcgcctc gccgcaccag gcggcagcag 3120

cgatacctgc cggtgccgcc tccgtcacgc gctcagccgc agctcagccc agccgcgagc 3180

tagggtttgt tcgtcctgaa ttgtttgatt tgatttgatt tgatttgatc cgatccgatc 3240

cgatctgatc tgatttgctt tgctttgctt tgtctccctc ccggcgcgga ccaagcgtcc 3300

gtctgcgcgc cgcagcttcc cttcttctcc cagccctcct tctgctcccg cctctcgcgc 3360

aagcacgcag cttcgccgcc gcatccggtc ggtcggtcgg tcgatcgacc cgcctgccgc 3420

tgctgctgtg gccgggcttt tctccatcgg cgactctttc ttctccatac gtcctactac 3480

gtacatacat actgccggct tcctcctctt ccagcgcggc gacggcggca ggctgcgacg 3540

tcgtcgccgc cgcgggcgcc gcgcgcgccg ccgccgccgc ccgcgtcgca gggcctcgtc 3600

gccgccgccg ctccgctccg ctccgaggcc gcgagagggc cgcggcggcg cgatggatgg 3660

atggatggat ggatggatgg atggattttg ttgatcgatg gcggcgcatg ggcggagatg 3720

agcgaggacg agcgcgcgag cgcggcagcc ggattcgcag ggcctcgctc gcctcgcgcc 3780

cgctgccgcg cccgccttgc gagcctgcgc cgcgagcgag cgagcgagcg agcggggctt 3840

tctttgtctc gcgcgccgct tggcctcgtg tgtcttgtgc ttgcgtagcg ggcgccgcgg 3900

tggaagatgg ctcattcaat cgacccattc acgcacgcac tccggcgcgc agagaaggcc 3960

gaggaggagc agcaagcaaa ccaaaagctc tcgcgctcgc ggtctcgggc tcgagcggtc 4020

tcggagagag agtcttgcgg cgaccaccgg cagcagcagc agcagcagca gcgctgtcga 4080

gcacgagcac gagcacgagc acgagcacga gcattcgagc aagaggacag acacggttgt 4140

cagcgcctag ctcgctcgat acagaaagag gcgggttggg cgtaaaaaaa aaggagcacg 4200

<210> 77
<211> 3886
<212> DNA
<213> Schizochytrium sp.

<220>
<221> misc_feature
<222> (2115)..(2115)
<223> n = a, c, g, or t

<400> 77
gatcttgatt gccaagctct ggattgtcga ttccgatgaa tcgagctctt tgttgtcgag 60

ctctggcttg ccgagctttc agaaatagac aaaattgccg agttcctgat tgcggggctc 120

tcgattgcca aggtctggtg gattctcgaa ctctcgattg tcaaaatctt ggtcgtctcg 180

tcggattctt tcctgatttg ttttgtcaag accttgagat tgtgcaaaac cttgatcgtt 240

gacaaaccct tgatcgacag cagcctttca tcacgctcag ctcttgtcat tgattatatt 300

ccccctgaca gccaacacct tgatgcaggg tctcaacctt gatttttgga ggccatcatc 360

agcatcacgc cccggcactc accctcaaca ttcgacagcc aacgcttttt tttcttcgac 420

taggatctga gaataaaagc aggtcaccac gaccgtaggc caacgcgaca accatggaaa 480

taaagtgaca acgaacgact tgcaagttta aatgtaaaga gcagcaattg cccgcccaca 540

gacaaatgaa agcaggcgcc gagtcttatt tgaggaggtg ggcctgtggc aatgggcgaa 600

agaaaatcaa ggacaaggag agcaggttac gtaccggtat actggtatac gtacatggat 660

ggttcttggc aagttgacgg gatgtgtgcg agtgaccgtg gtagttaacg aaagagccgc 720

aagggcaagg aaagcaagag aatgcagact tttccacagg atggatgggt ccgcagcttg 780

ccgcatgatg aaacgctgta tttcacctgg cacgtggtgg cgcacgcgcc cacatatgat 840

cgcggcggcg ggtgtattat acattttccc cctcaggtct actgccatcc ctccatgcgt 900

cgctcgtgcg aacgacgcaa gcctttcgca tcgtgcagcc tctttctggt aaggcaagag 960

ctaaacccaa acctaaacga aagaacattt ttacctctct ctctctccca ttggtcgcgt 1020

gcgctccgcc gctcgctcct cctcctgcca gtgtcgcgcc ctaacttccc ccctccctcc 1080

ctccctccct ccctccctct ctcctgccac cgcccctctc tccgcgctgc gtgcggtgct 1140

gccctggacc aatggcatgc tgctgcacgc tcggcggatg acgcaagccg cttcgcaatt 1200

tccggatcag atctcggcgg ggcgtgcgcc gcggggtcac tgcggacctg ccgcggcccc 1260

tgcttctttc acatccatca tgtcctccaa acctccgcct cctccacgca cgtacgcacg 1320

cccgctcgca cgcgcgcact gccgctgcga aagcaagcgc ccgcccgccg cccggcgacg 1380

ggaaggcggc cgcggtctcc ctccgcggtt gcctcgctcc cgcgcggggc tgggcgggca 1440

gcagaaggcg ggtggcggcg gcggcttccg tcttcgtcag cggcctacgt cggcggcggc 1500

gcgcgagact acgcatgccc ttgcgtcatg cgctcgcagg tagccgccgc gggcctagcg 1560

tttccgctgg cgccgcgcct aagcccccgg cgcgcacggt attgccgcga taccgtacgg 1620

ccaagaccgc cgcagacgtc ggccctctcg cggccagcca gccagcagcg cagcggagga 1680

agagcgcgca ggcgcggcgg gagggcggcc gcggagcagc gcagagcggg gcggagcagc 1740

gcggagcaga acgggcagac tcggagcggg cagggcgggc agagctttgg ggtttaagga 1800

ccgggttacc ggcgaagtga gcggctgcgg ggagcggctg tgggaggggt gagtacgcaa 1860

gcacgatgcg agcgagagag agacgctgcc gcgaatcaag aaggtaggcg cgctgcgagg 1920

cgcggcggcg gagcggagcg agggagaggg agagggagag agagggaggg agacgtcgcc 1980

gcggcggggc ctggcctggc ctggtttggc ttggtcagcg cggccttgtc cgagcgtgca 2040

gctggagttg ggtggattca tttggatttt cttttgtttt tgtttttctc tctttcccgg 2100

aaagtgttgg ccggncggtg ttctttgttt tgatttcttc aaaagttttg gtggttggtt 2160

ctctctcttg gctctctgtc aggcggtccg gtccacgccc cggcctctcc tctcctctcc 2220

tctcctctcc tctccgtgcg tatacgtacg tacgtttgta tacgtacata catcccgccc 2280

gccgtgccgg cgagggtttg ctcagcctgg agcaatgcga tgcgatgcga tgcgatgcga 2340

cgcgacgcga cgcgagtcac tggttcgcgc tgtggctgtg gcttgcttgc ttacttgctt 2400

tcgagctctc ccgctttctt ctttccttct cacgccacca ccaacgaaag aagatcggcc 2460

ccggcacgcc gctgagaagg gctggcggcg atgacggcac gcgcgcccgc tgccacgttg 2520

gcgctcgctg ctgctgctgc tgctgctgct gctgctgctg ctgctgctgc tgctgcttct 2580

gcgcgcaggc tttgccacga ggccggcgtg ctggccgctg ccgcttccag tccgcgtgga 2640

gagatcgaat gagagataaa ctggatggat tcatcgaggg atgaatgaac gatggttgga 2700

tgccttttc ctttttcagg tccacagcgg gaagcaggag cgcgtgaatc tgccgccatc 2760

cgcatacgtc tgcatcgcat cgcatcgcat gcacgcatcg ctcgccggga gccacagacg 2820

ggcgacaggg cggccagcca gccaggcagc cagccaggca ggcaccagag ggccagagag 2880

cgcgcctcac gcacgcgccg cagtgcgcgc atcgctcgca gtgcagacct tgattccccg 2940

cgcggatctc cgcgagcccg aaacgaagag cgccgtacgg gcccatccta gcgtcgcctc 3000

gcaccgcatc gcatcgcatc gcgttcccta gagagtagta ctcgacgaag gcaccatttc 3060

cgcgctcctc ttcggcgcga tcgaggcccc cggcgccgcg acgatcgcgg cggccgcggc 3120

gctggcggcg gccctggcgc tcgcgctggc ggccgccgcg ggcgtctggc cctggcgcgc 3180

gcgggcgccg caggaggagc ggcagcggct gctcgccgcc agagaagagc gcgccgggcc 3240

cggggaggga cggggaggag aaggagaagg cgcgcaaggc ggccccgaaa gagaagaccc 3300

tggacttgaa cgcgaagaag aagaagaagg agaagaagtt gaagaagaag aagaagaagg 3360

agaggaagtt gaagaagacg aggagcaggc gcgttccaag gcgcgttctc ttccggaggc 3420

gcgttccagc tgcggcggcg gggcgggctg cggggcgggc gcgggcgcgg gtgcgggcag 3480

aggggacgcg cgcgcggagg cggagggggc cgagcgggag cccctgctgc tgcggggcgc 3540

ccgggccgca ggtgtggcgc gcgcgacgac ggaggcgacg acgccagcgg ccgcgacgac 3600

aaggccggcg gcgtcggcgg gcggaaggcc ccgcgcggag caggggcggg agcaggacaa 3660

ggcgcaggag caggagcagg gccgggagcg ggagcgggag cgggcggcgg agcccgaggc 3720

agaacccaat cgagatccag agcgagcaga ggccggccgc gagcccgagc ccgcgccgca 3780

gatcactagt accgctgcgg aatcacagca gcagcagcag cagcagcagc agcagcagca 3840

gcagcagcag ccacgagagg gagataaaga aaaagcggca gagacg 3886

<210> 78
<211> 20
<212> DNA
<213> Artificial

<220>

<223> synthetic

<400> 78
catatggcgc tccgtgtcaa 20

<210> 79
<211> 35
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 79
gccaggaagc tttgacatgg ggtgccagga catct 35

<210> 80
<211> 37
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 80
tcctggcacc ccatgtcaaa gcttcctggc aacccta 37

<210> 81
<211> 20
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 81
agtatacaga ggtgctgaca 20

<210> 82
<211> 28
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 82
gcaccccatg agcaagctcc ccggcaac 28

<210> 83
<211> 28
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 83
gtatacagag gcgcagacac gttgtaag 28

<210> 84
<211> 28

<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 84
ctgcagccag atgctcaaga tgtacatg 28

<210> 85
<211> 31
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 85
ggagcttgct catggggtgc caggacatct c 31

<210> 86
<211> 10
<212> DNA
<213> artificial

<220>
<223> synthetic

<400> 86
ggatccatgg 10

<210> 87
<211> 31
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 87
aagcttgtgc agtcaagtgc gcaaaaccat g 31

<210> 88
<211> 15
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 88
taacccgggt ctaga 15

<210> 89
<211> 31
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 89
gactgctagc ttaagcaagg attttcttaa c 31

<210> 90
<211> 31
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 90
gactggatcc tcctgatgcg gtattttctc c 31

<210> 91
<211> 12
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 91
ctaggcggcc gc 12

<210> 92
<211> 31
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 92
gactactagt ctaagaaacc attattatca t 31

<210> 93
<211> 31
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 93
gactggatcc agctttaaat aatcggtgtc a 31

<210> 94
<211> 46
<212> DNA
<213> Artificial

<220>
<223> synthetic

<400> 94
gactggatcc gccaccatgt tgcagcatac ttggctacca aaaccc 46

<210> 95
<211> 35
<212> DNA
<213> Artificial

<220>

<223> synthetic

<400> 95
gacttctaga tcaataatgc cagaattttg gctgc 35