Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2673954 C 2015/09/15

(11)(21) 2 673 954

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2007/01/25
(87) Date publication PCT/PCT Publication Date: 2008/07/31
(45) Date de délivrance/Issue Date: 2015/09/15
(85) Entrée phase nationale/National Entry: 2009/06/26
(86) N° demande PCT/PCT Application No.: IB 2007/000558
(87) N° publication PCT/PCT Publication No.: 2008/090395

(51) Cl.Int./Int.Cl. *C12N 9/18* (2006.01), *C12N 1/21* (2006.01), *C12N 15/75* (2006.01)

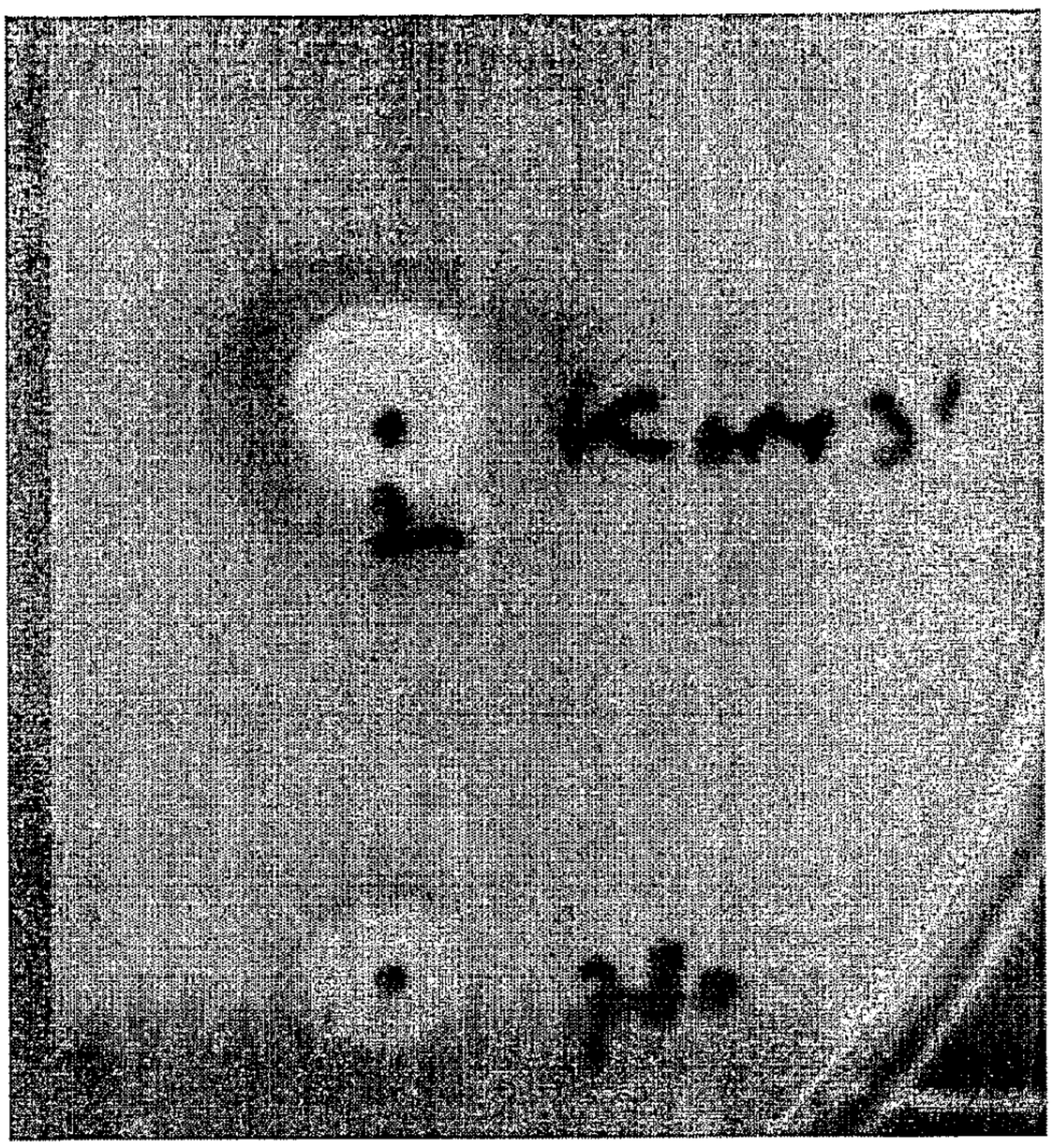
(72) Inventeurs/Inventors:
KOLKMAN, MARC, NL;
MIKKELSEN, JORN DALGAARD, DK;
LORENTSEN, RIKKE HOEGH, DK

(73) Propriétaire/Owner:
DUPONT NUTRITION BIOSCIENCES APS, DK

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : PRODUCTION D'UNE LIPIDE ACYLTRANSFERASE A PARTIR DE CELLULES TRANSFORMEES DE BACILLUS LICHENIFORMIS

(54) Title: PRODUCTION OF A LIPID ACYLTRANSFERASE FROM TRANSFORMED BACILLUS LICHENIFORMIS CELLS



(57) Abrégé/Abstract:
The present invention relates to a method for the production of a lipid acyltransferase comprising the steps of: (i) providing a Bacillus licheniformis cell; (ii) transforming the Bacillus licheniformis cell with a heterologous nucleotide sequence encoding a lipid acyltransferase; and (iii) expressing the lipid acyltransferase in the cell under the control of a promoter sequence. In addition, the present invention further relates to the use of Bacillus licheniformis to express a lipid acyltransferase, a Bacillus licheniformis host cell comprising a heterologous lipid acyltransferase and a vector comprising a nucleotide sequence encoding a lipid acyltransferase operably linked to a promoter sequence homologous to B. licheniformis.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 July 2008 (31.07.2008)

PCT

(10) International Publication Number
WO 2008/090395 A1

(51) International Patent Classification:
C12N 9/18 (2006.01) *C12N 1/21* (2006.01)
C12N 15/75 (2006.01)

(21) International Application Number:
PCT/IB2007/000558

(22) International Filing Date: 25 January 2007 (25.01.2007)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **DANISCO A/S** [DK/DK]; Langebrogade 1, P.O. Box 17, DK-1001 Copenhagen K (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KOLKMAN, Marc** [NL/NL]; Linnaeushof 25, NL-2341 PA Oegstgeest (NL). **MIKKELSEN, Jørn, Dalgaard** [DK/DK]; DK-2650 Hvidovre (DK). **LORENTSEN, Rikke, Høegh** [DK/DK]; Karetmagervej 52, DK-8900 Randers (DK).

(74) Agents: **WILLIAMS, Aylsa** et al.; D Young & Co., 120 Holborn, London EC1N 2DY (GB).

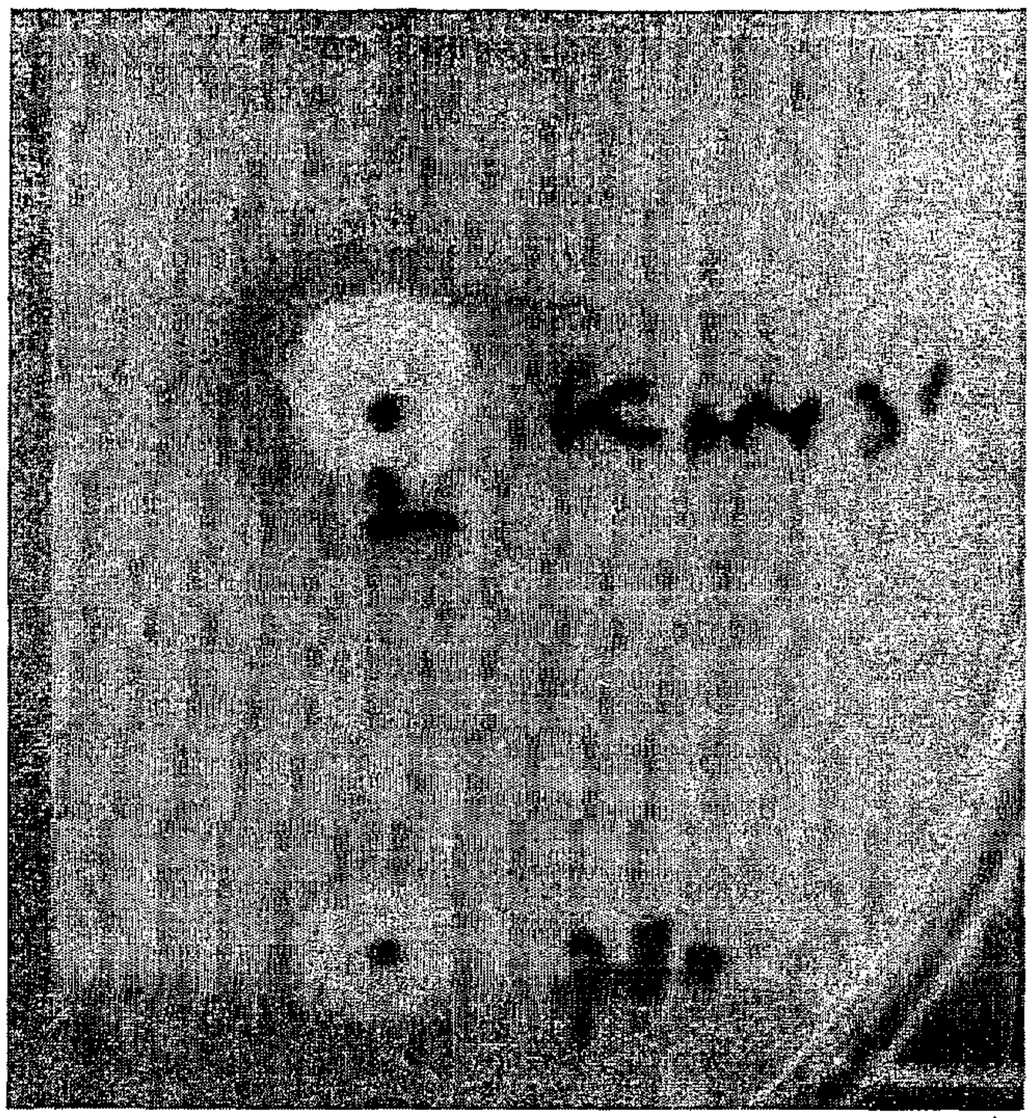
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR,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

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

Published:
— *with international search report*

(54) Title: PRODUCTION OF A LIPID ACYLTRANSFERASE FROM TRANSFORMED BACILLUS LICHENIFORMIS CELLS

FIGURE 56



(57) **Abstract:** The present invention relates to a method for the production of a lipid acyltransferase comprising the steps of: (i) providing a *Bacillus licheniformis* cell; (ii) transforming the *Bacillus licheniformis* cell with a heterologous nucleotide sequence encoding a lipid acyltransferase; and (iii) expressing the lipid acyltransferase in the cell under the control of a promoter sequence. In addition, the present invention further relates to the use of *Bacillus licheniformis* to express a lipid acyltransferase, a *Bacillus licheniformis* host cell comprising a heterologous lipid acyltransferase and a vector comprising a nucleotide sequence encoding a lipid acyltransferase operably linked to a promoter sequence homologous to *B. licheniformis*.

WO 2008/090395 A1

PRODUCTION OF A LIPID ACYLTRANSFERASE FROM TRANSFORMED BACILLUS LICHENIFORMIS CELLS

FIELD OF THE PRESENT INVENTION

The present invention relates to the production of lipid acyltransferases. In particular, methods for the production of a lipid acyltransferase by expressing a lipid acyltransferase in a *Bacillus* host cell, preferably a *B. licheniformis* host cell. In addition, the present invention relates to the use of *Bacillus* (preferably *B. licheniformis)* to express a lipid acyltransferase and to a *Bacillus* host cell, preferably a *B. licheniformis* host cell, comprising in its genome a gene encoding a lipid acyltransferase.

BACKGROUND OF THE PRESENT INVENTION

Lipid acyltransferases are known to be advantageous in food applications. Lipid acyltransferases have been found to have significant acyltransferase activity in foodstuffs. This activity has surprising beneficial applications in methods of preparing foodstuffs.

For instance, WO 2004/064537 discloses a method for the *in* situ production of an emulsifier by use of a lipid acyltransferase and the advantages associated therewith.

Accordingly, there is a need for a method for the commercial production of lipid acyltransferases.

However, generally genes can be difficult to express in heterologous hosts and expression of lipid acyltransferases in host cells can be problematic.

WO 2004/064537 discloses the expression of two *Aeromonas* lipid acyltransferases in *Bacillus subtilis* and *Escherichia Coli*. However, expression in *B. subtilis* is low whilst *E. coli* is not a GRAS organism and is, therefore, unsuitable as a host for enzymes that are to be used in the food industry.

US 6,255,076 discloses a method of producing a polypeptide in a *Bacillus* host cell. However, such a method requires the use of a tandem promoter in which each promoter sequence in operably linked to a single copy of a nucleic acid sequence encoding the polypeptide sequence. Thus, there is a need in the art for an improved method for the production of lipid acyltransferases.

SUMMARY ASPECTS OF THE PRESENT INVENTION

Aspects of the present invention are presented in the claims and in the following commentary.

One aspect of the present invention relates to a method for the production of a lipid acyltransferase comprising the steps of:

(i) providing a host cell, preferably a *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably a *Bacillus licheniformis* cell;

(ii) transforming the host cell, preferably the *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably the *Bacillus licheniformis* cell, with a heterologous nucleotide sequence encoding a lipid acyltransferase and

(iii) expressing the lipid acyltransferase in the cell under the control of a promoter sequence.

In another aspect, the present invention relates to a *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably a *Bacillus licheniformis* host cell, comprising a heterologous lipid acyltransferase.

In a further aspect, the present invention relates to the use of a *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably a *Bacillus licheniformis* host cell, in the production of a heterologous lipid acyltransferase.

Suitably expression in the *Bacillus* host wherein the *Bacillus* host is one other than *Bacillus subtilis*, and preferably wherein the *Bacillus* host is *B. licheniformis,* may result in increased expression when compared to expression in *B. subtilis*.

In yet another aspect, the present invention relates to an expression vector comprising a nucleotide sequence encoding a lipid acyltransferase operably linked to one or more regulatory sequence(s) such that the regulatory sequence(s) is capable of expressing the nucleotide sequence encoding a lipid acyltransferase in a suitable host or host cell, preferably in a *Bacillus* host (or cell) wherein the *Bacillus* host (or cell) is one other than *Bacillus subtilis*, preferably in *B. licheniformis* or a *B. licheniformis* cell.

Suitably the lipid acyltransferase may be a recombinant lipid acyltransferase.

DETAILED ASPECTS OF THE PRESENT INVENTION

According to a first aspect of the present invention there is provided a method for the production of a lipid acyltransferase comprising the steps of:

(i) providing a host cell, preferably a *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably a *Bacillus licheniformis* cell;

(ii) transforming the host cell, preferably a *Bacillus* host cell wherein the *Bacillus* host cell is one other than *Bacillus subtilis*, preferably a *Bacillus licheniformis* cell, with an heterologous nucleotide sequence encoding a lipid acyltransferase; and

(iii) expressing the lipid acyltransferase in the cell under the control of a promoter sequence.

Additionally, a nucleotide sequence encoding a signal peptide may be operably linked to said heterologous nucleotide sequence encoding a lipid acyltransferase.

Suitably the method of the present invention may further comprise the additional step of isolating/recovering the lipid acyltransferase.

In another aspect, the present invention relates to a *Bacillus licheniformis* host cell comprising a heterologous lipid acyltransferase.

Suitably the lipid acyltransferase may be a recombinant lipid acyltransferase.

Suitably the promoter sequence used in accordance with the host cells, vectors, methods and/or uses of the present invention may be homologous to the host cell. "Homologous to the host cell" means originating within the host organism; i.e. a promoter sequence which is found naturally in the host organism. Suitably, the promoter sequence may be selected from the group consisting of a nucleotide sequence encoding: an α-amylase promoter, a protease promoter, a subtilisin promoter, a glutamic acid-specific protease promoter and a levansucrase promoter. Suitably the promoter sequence may be a nucleotide sequence encoding: the LAT (e.g. the alpha-amylase promoter from *B. licheniformis*, also known as AmyL), AprL (e.g. subtilisin Carlsberg promoter), EndoGluC (e.g. the glutamic-acid specific promoter from *B. licheniformis*), AmyQ (e.g. the alpha amylase promoter from *B. amyloliquefaciens* alpha-amylase promoter) and SacB (e.g. the *B. subtilis* levansucrase promoter).

In one embodiment of the present invention the promoter sequence is the -35 to -10 sequence of an alpha amylase promoter, preferably the -35 to -10 sequence of a *B. licheniformis* α-amylase promoter. The "-35 to -10 sequence" describes the position relative to the transcription start site. Both the "-35" and the "-10" are boxes, i.e. a number of nucleotides, each comprising 6 nucleotides and these boxes are separated by 17 nucleotides. These 17 nucleotides are often referred to as a "spacer". This is illustrated in Figure 55, where the -35 and the -10 boxes are underlined. For the avoidance of doubt, where "-35 to -10 sequence" is used herein it refers to a sequence from the start of the -35 box to the end of the -10 box i.e. including both the -35 box, the 17 nucleotide long spacer and the -10 box.

In some aspects, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and/or uses of the present invention may comprise a GDSx motif and/ or a GANDY motif.

Preferably, the lipid acyltransferase enzyme is characterised as an enzyme which possesses acyltransferase activity and which comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S.

Suitably, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the vectors, host cells, methods and/or uses of the present invention may be obtainable, preferably obtained, from an organism from one or more of the following genera: *Aeromonas*, *Streptomyces, Saccharomyces, Lactococcus, Mycobacterium, Streptococcus, Lactobacillus, Desulfitobacterium, Bacillus, Campylobacter, Vibrionaceae, Xylella, Sulfolobus, Aspergillus, Schizosaccharomyces, Listeria, Neisseria, Mesorhizobium, Ralstonia, Xanthomonas and Candida*. Preferably, the lipid acyltransferase is obtainable, preferably obtained, from an organism from the genus *Aeromonas.*

In some aspects of the present invention, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the vectors, host cells, methods and/or uses of the present invention encodes a lipid acyltransferase that comprises an aspartic acid residue at a position corresponding to N-80 in the amino acid sequence of the *Aeromonas hydrophila* lipid acyltransferase shown as SEQ ID No. 35.

In addition or in the alternative, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the vectors, host cells, methods and/or uses of the present invention encodes a lipid acyltransferase that may comprise the amino acid sequence shown as SEQ ID No. 16, or an amino acid sequence which has 75% or more homology thereto. Suitably, the nucleotide sequence encoding a lipid acyltransferase encodes a lipid acyltransferase that may comprise the amino acid sequence shown as SEQ ID No. 16.

The term "heterologous" as used herein means a sequence derived from a separate genetic source or species. A heterologous sequence is a non-host sequence, a modified

sequence, a sequence from a different host cell strain, or a homologous sequence from a different chromosomal location of the host cell.

A "homologous" sequence is a sequence that is found in the same genetic source or species i.e. it is naturally occurring in the relevant species of host cell.

The term "recombinant lipid acyltransferase" as used herein means that the lipid acyltransferase has been produced by means of genetic recombination. For instance, the nucleotide sequence encoding the lipid acyltansferase has been inserted into a cloning vector, resulting in a *B. licheniformis* cell characterised by the presence of the heterologous lipid acyltransferase.

HOST CELL

In one embodiment of the present invention the host cell for use in the methods and/or uses of thee present invention is a *Bacillus licheniformis* host cell.

It has been found that the use of a *Bacillus licheniformis* host cell results in increased expression of a lipid acyltransferase when compared with other organisms, such as *Bacillus subtilis*.

A lipid acyltransferase from *Aeromonas salmonicid*a has been inserted into a number of conventional expression vectors, designed to be optimal for the expression in *Bacillus subtilis*, *Hansenula polymorpha*, *Schizosaccharomyces pombe* and *Aspergillus tubigensis*, respectively. Only very low levels were, however, detected in *Hansenula polymorpha*, *Schizosaccharomyces pombe* and *Aspergillus tubigensis*. The expression levels were below 1 µg/ml, and it was not possible to select cells which yielded enough protein to initiate a commercial production (results not shown). In contrast, *Bacillus licheniformis* was able to produce protein levels, which are attractive for an economically feasible production.

In particular, it has been found that expression in *B. licheniformis* is approximately 100-times greater than expression in *B. subtilis* under the control of aprE promoter or is approximately 100-times greater than expression in *S. lividans* under the control of an A4 promoter and fused to cellulose (results not shown herein).

In another embodiment the host cell may be any *Bacillus* cell other than *B.subtilis.* Preferably, said *Bacillus* host cell being from one of the following species: *Bacillus licheniformis; B. alkalophilus; B. amyloliquefaciens; B. circulans; B. clausii; B. coagulans; B. firmus; B. lautus; B. lentus; B. megaterium; B. pumilus* or *B. stearothermophilus.*

The term "host cell" - in relation to the present invention includes any cell that comprises either a nucleotide sequence encoding a lipid acyltransferase as defined herein or an expression vector as described above and which is used in the recombinant production of a lipid acyltransferase having the specific properties as defined herein.

Thus, a further embodiment of the present invention provides a host cell comprising (for example transformed or transfected with) a nucleotide sequence of the present invention or a nucleotide sequence that expresses a polypeptide having the specific properties as defined herein.

Suitably, in some embodiments, the host cell may be a protease deficient or protease minus strain and/or an α-amylase deficient or α-amylase minus strain.

REGULATORY SEQUENCES

In some applications, a lipid acyltransferase sequence for use in any one of the host cells, vectors, methods and/or uses of the present invention may be operably linked to a regulatory sequence which is capable of providing for the expression of the nucleotide sequence, such as by the chosen host cell (such as a *B. licheniformis* cell).

By way of example, the present invention covers a vector comprising the nucleotide sequence of the present invention operably linked to such a regulatory sequence, i.e. the vector is an expression vector.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way

that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "regulatory sequences" includes promoters and enhancers and other expression regulation signals.

The term "promoter" is used in the normal sense of the art, e.g. an RNA polymerase binding site.

Enhanced expression of the nucleotide sequence encoding the enzyme having the specific properties as defined herein may also be achieved by the selection of regulatory regions, e.g. promoter, secretion leader and terminator regions that are not regulatory regions for the nucleotide sequence encoding the enzyme in nature.

Suitably, the nucleotide sequence of the present invention may be operably linked to at least a promoter.

Suitably, the nucleotide sequence encoding a lipid acyltransferase may be operably linked to at a nucleotide sequence encoding a terminator sequence. Examples of suitable terminator sequences for use in any one of the vectors, host cells, methods and/or uses of the present invention include: an α-amylase terminator sequence (for instance, CGGGACTTACCGAAAGAAACCATCAATGATGGTTTCTTTTTTGTTCATAAA – SEQ ID No. 64), an alkaline protease terminator sequence (for instance, CAAGACTAAAGACCGTTCGCCCGTTTTTGCAATAAGCGGGCGAATCTTACATAAAA ATA – SEQ ID No. 65), a glutamic-acid specific terminator sequence (for instance, ACGGCCGTTAGATGTGACAGCCCGTTCCAAAAGGAAGCGGGCTGTCTTCGTGTAT TATTGT – SEQ ID No. 66), a levanase terminator sequence (for instance, TCTTTTAAAGGAAAGGCTGGAATGCCCGGCATTCCAGCCACATGATCATCGTTT – SEQ ID No. 67) and a subtilisin E terminator sequence (for instance, GCTGACAAATAAAAAGAAGCAGGTATGGAGGAACCTGCTTCTTTTTACTATTATTG). Suitably, the nucleotide sequence encoding a lipid acyltransferase may be operably linked to an α-amylase terminator, such as a *B. licheniformis* α-amylase terminator.

PROMOTER

The promoter sequence to be used in accordance with the present invention may be heterologous or homologous to the sequence encoding a lipid acyltransferase.

The promoter sequence may be any promoter sequence capable of directing expression of a lipid acyltransferase in the host cell of choice.

Suitably, the promoter sequence may be homologous to a *Bacillus* species, for example *B. licheniformis*. Preferably, the promoter sequence is homologous to the host cell of choice.

Suitable promoter sequences for use in the present invention include: the promoter of the *Bacillus licheniformis* alpha-amylase gene, the promoter of the *Bacillus licheniformis* subtilisin gene, the promoter of the *Bacillus subtilis* subtilisin gene, the promoter of the *Bacillus licheniformis* alkaline protease gene (subtilisin Carlsberg gene), the promoter of the *B. licheniformis* glutamic-acid specific protease gene, the promoter of *B. amyloliquefaciens* alpha-amylase gene; the promoter of *B. subtilis* levansucrase and a "consensus" promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region (i.e. the -35 to -10 promoter) of the alpha-amylase gene.

Other examples of promoters suitable for directing the transcription of a nucleic acid sequence in the methods of the present invention include: the promoter of the *Bacillus lentus* alkaline protease gene (aprH), ; the promoter of the *Bacillus subtilis* alpha-amylase gene (amyE); the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene (amyM); the promoter of the *Bacillus licheniformis* penicillinase gene (penP); the promoters of the *Bacillus subtilis* xylA and xylB genes; and/or the promoter of the *Bacillus thuringiensis* subsp. *tenebrionis* CryIIIA gene.

In a preferred embodiment, the promoter sequence is an α-amylase promoter (such as a *Bacillus licheniformis* α-amylase promoter). Preferably, the promoter sequence comprises the -35 to -10 sequence of the B. *licheniformis* α-amylase promoter – see Figures 53 and 55.

SIGNAL PEPTIDE

The lipid acyltransferase produced by a host cell by expression of the nucleotide sequence encoding the lipid acyltransferase may be secreted or may be contained intracellularly depending on the sequence and/or the vector used.

A signal sequence may be used to direct secretion of the coding sequences through a particular cell membrane. The signal sequences may be natural or foreign to the lipid acyltransferase coding sequence. For instance, the signal peptide coding sequence may be obtained form an amylase or protease gene from a *Bacillus* species, preferably from *Bacillus licheniformis.*

Suitable signal peptide coding sequences may be obtained from one or more of the following genes: maltogenic α-amylase gene, subtilisin gene, beta-lactamase gene, neutral protease gene, prsA gene, and/or acyltransferase gene.

Preferably, the signal peptide is a signal peptide of *B. licheniformis* α-amylase, Aeromonas acyltransferase (for instance, mkkwfvcllglialtvqa - SEQ ID No. 21), *B. subtilis* subtilisin (for instance, mrskklwisllfaltliftmafsnmsaqa - SEQ ID No. 22) or *B. licheniformis* subtilisin (for instance, mmrkksfwfgmltafmlvftmefsdsasa - SEQ ID No. 23). Suitably, the signal peptide may be the signal peptide of *B. licheniformis* α-amylase.

However, any signal peptide coding sequence capable of directing the expressed lipid acyltransferase into the secretory pathway of a *Bacillus host* cell (preferably a *B. licheniformis* host cell) of choice may be used.

In some embodiments of the present invention, a nucleotide sequence encoding a signal peptide may be operably linked to a nucleotide sequence encoding a lipid acyltransferase of choice.

The lipid acyltransferase of choice may be expressed in a host cell as defined herein as a fusion protein.

EXPRESSION VECTOR

The term "expression vector" means a construct capable of *in vivo* or *in vitro* expression.

Preferably, the expression vector is incorporated in the genome of the organism, such as a *B. licheniformis* host. The term "incorporated" preferably covers stable incorporation into the genome.

The nucleotide sequence encoding a lipid acyltransferase as defined herein may be present in a vector, in which the nucleotide sequence is operably linked to regulatory sequences such that the regulatory sequences are capable of providing the expression of the nucleotide sequence by a suitable host organism (such as *B. licheniformis*), i.e. the vector is an expression vector.

The vectors of the present invention may be transformed into a suitable host cell as described above to provide for expression of a polypeptide having lipid acyltransferase activity as defined herein.

The choice of vector, e.g. plasmid, cosmid, virus or phage vector, genomic insert, will often depend on the host cell into which it is to be introduced. The present invention may cover other forms of expression vectors which serve equivalent functions and which are, or become, known in the art.

Once transformed into the host cell of choice, the vector may replicate and function independently of the host cell's genome, or may integrate into the genome itself.

The vectors may contain one or more selectable marker genes – such as a gene which confers antibiotic resistance e.g. ampicillin, kanamycin, chloramphenicol or tetracyclin resistance. Alternatively, the selection may be accomplished by co-transformation (as described in WO91/17243).

Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell.

Thus, in a further embodiment, the invention provides a method of making nucleotide sequences of the present invention or nucleotide sequences encoding polypeptides having the specific properties as defined herein for use in any one of the vectors, host cells, other methods and/or uses of the present invention, by introducing a nucleotide sequence into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector.

The vector may further comprise a nucleotide sequence enabling the vector to replicate in the host cell in question. Examples of such sequences are the origins of replication of plasmids pUC19, pACYC177, pUB110, pE194, pAMB1 and pIJ702.

LIPID ACYL TRANSFERASE

The nucleotide sequence encoding a lipid acyl transferase for use in any one of the methods, vectors and/or uses of the present invention may encode a natural lipid acyl transferase or a variant lipid acyl transferase.

For instance, the nucleotide sequence encoding a lipid acyl transferase for use in the present invention may be one as described in WO2004/064537, WO2004/064987, WO2005/066347, or WO2006/008508.

The term "lipid acyl transferase" as used herein preferably means an enzyme that has acyltransferase activity (generally classified as E.C. 2.3.1.x, for example 2.3.1.43), whereby the enzyme is capable of transferring an acyl group from a lipid to one or more acceptor substrates, such as one or more of the following: a sterol; a stanol; a carbohydrate; a protein; a protein subunit; a sugar alcohol, such as ascorbic acid and/or glycerol – preferably glycerol and/or a sterol, such as cholesterol.

Preferably, the nucleotide sequence encoding a lipid acyl transferase for use in any one of the vectors, host cells, methods and/or uses of the present invention encodes a lipid acyltransferase that is capable of transferring an acyl group from a

phospholipid (as defined herein) to a sugar alcohol, such as ascorbic acid and/or glycerol, most preferably glycerol.

For some aspects the "acyl acceptor" according to the present invention may be any compound comprising a hydroxy group (-OH), such as for example, polyvalent alcohols, including glycerol; sterols; stanols; carbohydrates; hydroxy acids including fruit acids, citric acid, tartaric acid, lactic acid and ascorbic acid; proteins or a sub-unit thereof, such as amino acids, protein hydrolysates and peptides (partly hydrolysed protein) for example; and mixtures and derivatives thereof. Preferably, the "acyl acceptor" according to the present invention is not water. Preferably, the "acyl acceptor" according to the present invention is a sugar alcohol, such as a polyol, most preferably glycerol. For the purpose of this invention ascorbic acid is also considered a sugar-alcohol.

The acyl acceptor is preferably not a monoglyceride.

The acyl acceptor is preferably not a diglyceride

In one aspect, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that may, as well as being able to transfer an acyl group from a lipid to glycerol, additionally be able to transfer the acyl group from a lipid to one or more of the following: a carbohydrate, a protein, a protein subunit, sterol and/or a stanol, preferably it is capable of transferring to both a sugar alcohol, such as ascorbic acid and/or glycerol, most preferably a sterol such as cholesterol, and/or plant sterol/stanols.

Preferably, the lipid substrate upon which the lipid acyl acts is one or more of the following lipids: a phospholipid, such as a lecithin, e.g. phosphatidylcholine.

This lipid substrate may be referred to herein as the "lipid acyl donor". The term lecithin as used herein encompasses phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylserine and phosphatidylglycerol.

For some aspects, preferably the nucleotide sequence encoding a lipid acyl transferase for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that is incapable, or substantially incapable, of acting on a triglyceride and/or a 1-monoglyceride and/or 2-monoglyceride.

For some aspects, preferably the nucleotide sequence encoding a lipid acyl transferase for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that does not exhibit triacylglycerol lipase activity (E.C. 3.1.1.3) or does not exhibit significant triacylglycerol lipase activity (E.C. 3.1.1.3).

The ability to hydrolyse triglyeride (E.C. 3.1.1.3 activity) may be determined by lipase activity is determined according to Food Chemical Codex (3rd Ed., 1981, pp 492-493) modified to sunflower oil and pH 5.5 instead of olive oil and pH 6.5. The lipase activity is measured as LUS (lipase units sunflower) where 1 LUS is defined as the quantity of enzyme which can release 1 [mu]mol of fatty acids per minute from sunflower oil under the above assay conditions. Alternatively the LUT assay as defined in WO9845453 may be used.

The nucleotide sequence encoding a lipid acyl transferase for use in any one of the host cells, vectors, methods and/or uses of the present invention may encode a lipid acyltransferase that which is substantially incapable of acting on a triglyceride may have a LUS/mg of less than 1000, for example less than 500, such as less than 300, preferably less than 200, more preferably less than 100, more preferably less than 50, more preferably less than 20, more preferably less than 10, such as less than 5, less than 2, more preferably less than 1 LUS/mg. Alternatively LUT/mg activity is less than 500, such as less than 300, preferably less than 200, more preferably less than 100, more preferably less than 50, more preferably less than 20, more preferably less than 10, such as less than 5, less than 2, more preferably less than 1 LUT/mg.

The nucleotide sequence encoding a lipid acyl transferase for use in any one of the host cells, vectors, methods and/or uses of the present invention may encode a lipid acyltransferase that which is substantially incapable of acting on a monoglyceride may be determined by using mono-oleate (M7765 1-Oleoyl-*rac*-glycerol 99%) in place

of the sunflower oil in the LUS assay. 1 MGHU is defined as the quantity of enzyme which can release 1 [mu]mol of fatty acids per minute from monoglyceride under the assay conditions.

The nucleotide sequence encoding a lipid acyl transferase for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that which is substantially incapable of acting on a triglyceride may have a MGHU/mg of less than 5000, for example less than 1000, for example less than 500, such as less than 300, preferably less than 200, more preferably less than 100, more preferably less than 50, more preferably less than 20, more preferably less than 10, such as less than 5, less than 2, more preferably less than 1 MGHU/mg.

Suitably, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that may exhibit one or more of the following phospholipase activities: phospholipase A2 activity (E.C. 3.1.1.4) and/or phospholipase A1 activity (E.C. 3.1.1.32). The lipid acyl transferase may also have phospholipase B activity (E.C 3.1.1.5).

Suitably, for some aspects the lipid acyltransferase may be capable of transferring an acyl group from a phospholipid to a sugar alcohol, preferably glycerol and/or ascorbic acid.

For some aspects, preferably the nucleotide sequence encoding a lipid acyltransferase for use any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that is capable of transferring an acyl group from a phospholipid to a sterol and/or a stanol to form at least a sterol ester and/or a stanol ester.

The lipid acyltransferase may be capable of transferring an acyl group from a lipid to a polyol such as glycerol, and/or a sterol such as cholesterol or plant sterol/stanols. Thus, in one embodiment the “acyl acceptor” according to the present invention may be glycerol and/or cholesterol or plant sterol/stanols.

Preferably, the lipid acyltransferase enzyme may be characterised using the following criteria:

the enzyme possesses acyl transferase activity which may be defined as ester transfer activity whereby the acyl part of an original ester bond of a lipid acyl donor is transferred to an acyl acceptor, preferably glycerol or cholesterol, to form a new ester; and

the enzyme comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S.

Preferably, X of the GDSX motif is L or Y. More preferably, X of the GDSX motif is L. Thus, preferably the enzyme according to the present invention comprises the amino acid sequence motif GDSL.

The GDSX motif is comprised of four conserved amino acids. Preferably, the serine within the motif is a catalytic serine of the lipid acyl transferase enzyme. Suitably, the serine of the GDSX motif may be in a position corresponding to Ser-16 in *Aeromonas hydrophila* lipid acyltransferase enzyme taught in Brumlik & Buckley (Journal of Bacteriology Apr. 1996, Vol. 178, No. 7, p 2060-2064).

To determine if a protein has the GDSX motif according to the present invention, the sequence is preferably compared with the hidden markov model profiles (HMM profiles) of the pfam database in accordance with the procedures taught in WO2004/064537 or WO2004/064987, incorporated herein by reference.

Preferably the lipid acyl transferase enzyme can be aligned using the Pfam00657 consensus sequence (for a full explanation see WO2004/064537 or WO2004/064987).

Preferably, a positive match with the hidden markov model profile (HMM profile) of the pfam00657 domain family indicates the presence of the GDSL or GDSX domain according to the present invention.

Preferably when aligned with the Pfam00657 consensus sequence the lipid acyltransferase for use in the methods or uses of the invention may have at least

one, preferably more than one, preferably more than two, of the following, a GDSx block, a GANDY block, a HPT block. Suitably, the lipid acyltransferase may have a GDSx block and a GANDY block. Alternatively, the enzyme may have a GDSx block and a HPT block. Preferably the enzyme comprises at least a GDSx block. See WO2004/064537 or WO2004/064987 for further details.

Preferably, residues of the GANDY motif are selected from GANDY, GGNDA, GGNDL, most preferably GANDY.

Preferably, when aligned with the Pfam00657 consensus sequence the enzyme for use in the methods or uses of the invention have at least one, preferably more than one, preferably more than two, preferably more than three, preferably more than four, preferably more than five, preferably more than six, preferably more than seven, preferably more than eight, preferably more than nine, preferably more than ten, preferably more than eleven, preferably more than twelve, preferably more than thirteen, preferably more than fourteen, of the following amino acid residues when compared to the reference *A. hydrophilia* polypeptide sequence, namely SEQ ID No. 1: 28hid, 29hid, 30hid, 31hid, 32gly, 33Asp, 34Ser, 35hid, 130hid, 131Gly, 132Hid, 133Asn, 134Asp, 135hid, 309His.

The pfam00657 GDSX domain is a unique identifier which distinguishes proteins possessing this domain from other enzymes.

The pfam00657 consensus sequence is presented in Figure 3 as SEQ ID No. 2. This is derived from the identification of the pfam family 00657, database version 6, which may also be referred to as pfam00657.6 herein.

The consensus sequence may be updated by using further releases of the pfam database (for example see WO2004/064537 or WO2004/064987).

In one embodiment, the nucleotide sequence encoding a lipid acyl transferase enzyme for use in any one of the host cells, vectors, methods and/or uses of the present invention encodes a lipid acyltransferase that may be characterised using the following criteria:

(i) the enzyme possesses acyl transferase activity which may be defined as ester transfer activity whereby the acyl part of an original ester bond of a lipid acyl donor is transferred to acyl acceptor, preferably glycerol or cholesterol, to form a new ester, preferably monoglyceride or cholesterol ester respectfully;

(ii) the enzyme comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S.;

(iii) the enzyme comprises His-309 or comprises a histidine residue at a position corresponding to His-309 in the *Aeromonas hydrophila* lipid acyltransferase enzyme shown in Figures 2 and 4 (SEQ ID No. 1 or SEQ ID No. 3).

Preferably, the amino acid residue of the GDSX motif is L.

In SEQ ID No. 3 or SEQ ID No. 1 the first 18 amino acid residues form a signal sequence. His-309 of the full length sequence, that is the protein including the signal sequence, equates to His-291 of the mature part of the protein, i.e. the sequence without the signal sequence.

In one embodiment, the nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that comprises the following catalytic triad: Ser-34, Asp-306 and His-309 or comprises a serine residue, an aspartic acid residue and a histidine residue, respectively, at positions corresponding to Ser-34, Asp-306 and His-309 in the *Aeromonas hydrophila* lipid acyl transferase enzyme shown in Figure 4 (SEQ ID No. 3) or Figure 2 (SEQ ID No. 1). As stated above, in the sequence shown in SEQ ID No. 3 or SEQ ID No. 1 the first 18 amino acid residues form a signal sequence. Ser-34, Asp-306 and His-309 of the full length sequence, that is the protein including the signal sequence, equate to Ser-16, Asp-288 and His-291 of the mature part of the protein, i.e. the sequence without the signal sequence. In the pfam00657 consensus sequence, as given in Figure 3 (SEQ ID No. 2) the active site residues correspond to Ser-7, Asp-345 and His-348.

In one embodiment, the nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may be characterised using the following criteria:

the enzyme possesses acyl transferase activity which may be defined as ester transfer activity whereby the acyl part of an original ester bond of a first lipid acyl donor is transferred to an acyl acceptor to form a new ester; and

the enzyme comprises at least Gly-32, Asp-33, Ser-34, Asp-134 and His-309 or comprises glycine, aspartic acid, serine, aspartic acid and histidine residues at positions corresponding to Gly-32, Asp-33, Ser-34, Asp-306 and His-309, respectively, in the *Aeromonas hydrophila* lipid acyltransferase enzyme shown in SEQ ID No. 3 or SEQ ID No. 1.

Suitably, the nucleotide sequence encoding a lipid acyltransferase enzyme for use in any one of the host cells, vectors, methods and uses of the present invention may be one of the following nucleotide sequences:

(a) the nucleotide sequence shown as SEQ ID No. 36 (see Figure 29);
(b) the nucleotide sequence shown as SEQ ID No. 38 (see Figure 31);
(c) the nucleotide sequence shown as SEQ ID No. 39 (see Figure 32);
(d) the nucleotide sequence shown as SEQ ID No. 42 (see Figure 35);
(e) the nucleotide sequence shown as SEQ ID No. 44 (see Figure 37);
(f) the nucleotide sequence shown as SEQ ID No. 46 (see Figure 39);
(g) the nucleotide sequence shown as SEQ ID No. 48 (see Figure 41);
(h) the nucleotide sequence shown as SEQ ID No. 49 (see Figure 57);
(i) the nucleotide sequence shown as SEQ ID No. 50 (see Figure 58);
(j) the nucleotide sequence shown as SEQ ID No. 51 (see Figure 59);
(k) the nucleotide sequence shown as SEQ ID No. 52 (see Figure 60);
(l) the nucleotide sequence shown as SEQ ID No. 53 (see Figure 61);
(m) the nucleotide sequence shown as SEQ ID No. 54 (see Figure 62);
(n) the nucleotide sequence shown as SEQ ID No. 55 (see Figure 63);
(o) the nucleotide sequence shown as SEQ ID No. 56 (see Figure 64);
(p) the nucleotide sequence shown as SEQ ID No. 57 (see Figure 65);
(q) the nucleotide sequence shown as SEQ ID No. 58 (see Figure 66);
(r) the nucleotide sequence shown as SEQ ID No. 59 (see Figure 67);

(s) the nucleotide sequence shown as SEQ ID No. 60 (see Figure 68);

(t) the nucleotide sequence shown as SEQ ID No. 61 (see Figure 69);

(u) the nucleotide sequence shown as SEQ ID No. 62 (see Figure 70);

(v) the nucleotide sequence shown as SEQ ID No. 63 (see Figure 71);

(w) or

a nucleotide sequence which has 70% or more, preferably 75% or more, identity with any one of the sequences shown as SEQ ID No. 36, SEQ ID No. 38, SEQ ID No. 39, SEQ ID No. 42, SEQ ID No. 44, SEQ ID No. 46, SEQ ID No. 48, SEQ ID No. 49, SEQ ID No. 50, SEQ ID No. 51, SEQ ID No. 52, SEQ ID No. 53, SEQ ID No. 54, SEQ ID No. 55, SEQ ID No. 56, SEQ ID No. 57, SEQ ID No. 58, SEQ ID No. 59, SEQ ID No. 60, SEQ ID No. 61, SEQ ID No. 62 or SEQ ID No. 63.

Suitably the nucleotide sequence may have 80% or more, preferably 85% or more, more preferably 90% or more and even more preferably 95% or more identity with any one of the sequences shown as SEQ ID No. 36, SEQ ID No. 38, SEQ ID No. 39, SEQ ID No. 42, SEQ ID No. 44, SEQ ID No. 46, SEQ ID No. 48, SEQ ID No. 49, SEQ ID No. 50, SEQ ID No. 51, SEQ ID No. 52, SEQ ID No. 53, SEQ ID No. 54, SEQ ID No. 55, SEQ ID No. 56, SEQ ID No. 57, SEQ ID No. 58, SEQ ID No. 59, SEQ ID No. 60, SEQ ID No. 61, SEQ ID No. 62 or SEQ ID No. 63.

In one embodiment, the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention is a nucleotide sequence which has 70% or more, preferably 75% or more, identity with any one of the sequences shown as: SEQ ID No. 49, SEQ ID No. 50, SEQ ID No. 51, SEQ ID No. 62, and SEQ ID No. 63. Suitably the nucleotide sequence may have 80% or more, preferably 85% or more, more preferably 90% or more and even more preferably 95% or more identity with any one of the sequences shown as: SEQ ID No. 49, SEQ ID No. 50, SEQ ID No. 51, SEQ ID No. 62, and SEQ ID No. 63.

In one embodiment, the nucleotide sequence encoding a lipid acyltransferase enzyme for use in any one of the host cells, vectors, methods and uses of the present invention is a nucleotide sequence which has 70% or more, 75% or more, 80% or more, preferably 85% or more, more preferably 90% or more and even more preferably 95% or more identity the sequence shown as SEQ ID No. 49.

Suitably, the nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises one or more of the following amino acid sequences:

(i) the amino acid sequence shown as SEQ ID No. 3
(ii) the amino acid sequence shown as SEQ ID No. 4
(iii) the amino acid sequence shown as SEQ ID No. 5
(iv) the amino acid sequence shown as SEQ ID No. 6
(v) the amino acid sequence shown as SEQ ID No. 7
(vi) the amino acid sequence shown as SEQ ID No. 8
(vii) the amino acid sequence shown as SEQ ID No. 9
(viii) the amino acid sequence shown as SEQ ID No. 10
(ix) the amino acid sequence shown as SEQ ID No. 11
(x) the amino acid sequence shown as SEQ ID No. 12
(xi) the amino acid sequence shown as SEQ ID No. 13
(xii) the amino acid sequence shown as SEQ ID No. 14
(xiii) the amino acid sequence shown as SEQ ID No. 1
(xiv) the amino acid sequence shown as SEQ ID No. 15 or

an amino acid sequence which has 75%, 80%, 85%, 90%, 95%, 98% or more identity with any one of the sequences shown as SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, or SEQ ID No. 15.

Suitably, nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises either the amino acid sequence shown as SEQ ID No. 3 or as SEQ ID No. 4 or SEQ ID No. 1 or SEQ ID No. 15 or comprises an amino acid sequence which has 75% or more, preferably 80% or more, preferably 85% or more, preferably 90% or more, preferably 95% or more, identity with the amino acid sequence shown as SEQ ID No. 3 or the amino acid sequence shown as SEQ ID No. 4 or the amino acid sequence shown as SEQ ID No. 1 or the amino acid sequence shown as SEQ ID No. 15.

Suitably the nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises an amino acid sequence which has 80% or more, preferably 85% or more, more preferably 90% or more and even more preferably 95% or more identity with any one of the sequences shown as SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, or SEQ ID No. 15.

Suitably, the nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises one or more of the following amino acid sequences:

(a) an amino acid sequence shown as amino acid residues 1-100 of SEQ ID No. 3 or SEQ ID No. 1;

(b) an amino acid sequence shown as amino acids residues 101-200 of SEQ ID No. 3 or SEQ ID No. 1;

(c) an amino acid sequence shown as amino acid residues 201-300 of SEQ ID No. 3 or SEQ ID No. 1; or

(d) an amino acid sequence which has 75% or more, preferably 85% or more, more preferably 90% or more, even more preferably 95% or more identity to any one of the amino acid sequences defined in (a)-(c) above.

Suitably, lipid acyl transferase enzyme for use in methods and uses of the present invention may comprise one or more of the following amino acid sequences:

(a) an amino acid sequence shown as amino acid residues 28-39 of SEQ ID No. 3 or SEQ ID No. 1;

(b) an amino acid sequence shown as amino acids residues 77-88 of SEQ ID No. 3 or SEQ ID No. 1;

(c) an amino acid sequence shown as amino acid residues 126-136 of SEQ ID No. 3 or SEQ ID No. 1;

(d) an amino acid sequence shown as amino acid residues 163-175 of SEQ ID No. 3 or SEQ ID No. 1;

(e) an amino acid sequence shown as amino acid residues 304-311 of SEQ ID No. 3 or SEQ ID No. 1; or

(f) an amino acid sequence which has 75% or more, preferably 85% or more, more preferably 90% or more, even more preferably 95% or more identity to any one of the amino acid sequences defined in (a)-(e) above.

In one aspect, nucleotide sequence encoding a lipid acyl transferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may be the lipid acyl transferase from *Candida parapsilosis* as taught in EP 1 275 711. Thus in one aspect the lipid acyl transferase for use in the method and uses of the present invention may be a lipid acyl transferase comprising one of the amino acid sequences taught in SEQ ID No. 17 or SEQ ID No. 18.

Much by preference, the nucleotide sequence encoding a lipid acyl transferase enzyme for use in any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may be a lipid acyl transferase (lipid acyltransferase) comprising the amino acid sequence shown as SEQ ID No. 16, or an amino acid sequence which has 75% or more, preferably 85% or more, more preferably 90% or more, even more preferably 95% or more, even more preferably 98% or more, or even more preferably 99% or more identity to SEQ ID No. 16. This enzyme could be considered a variant enzyme.

In one aspect, the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may be a lecithin:cholesterol acyltransferase (LCAT) or variant thereof (for example a variant made by molecular evolution)

Suitable LCATs are known in the art and may be obtainable from one or more of the following organisms for example: mammals, rat, mice, chickens, *Drosophila melanogaster*, plants, including Arabidopsis and *Oryza sativa*, nematodes, fungi and yeast.

In one embodiment the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may be the lipid acyltransferase obtainable, preferably obtained, from the E. coli strains TOP 10 harbouring pPet12aAhydro and

pPet12aASalmo deposited by Danisco A/S of Langebrogade 1, DK-1001 Copenhagen K, Denmark under the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the purposes of Patent Procedure at the National Collection of Industrial, Marine and Food Bacteria (NCIMB) 23 St. Machar Street, Aberdeen Scotland, GB on 22 December 2003 under accession numbers NCIMB 41204 and NCIMB 41205, respectively.

A nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a phospholipid glycerol acyl transferase. Phospholipid glycerol acyl transferases include those isolated from *Aeromonas spp.*, preferably *Aeromonas hydrophila* or *A. salmonicida*, most preferable *A. salmonicida* or variants thereof. Most preferred lipid acyl transferases for use in the present invention are encoded by SEQ ID No.s 1, 3, 4, 15 and 16. It will be recognised by the skilled person that it is preferable that the signal peptides of the acyl transferase has been cleaved during expression of the transferase. The signal peptide of SEQ ID 1, 3, 4, 15 and 16 are amino acids 1-18. Therefore the most preferred regions are amino acids 19-335 for SEQ ID No. 1 and SEQ ID No. 3 (*A. hydrophilia*) and amino acids 19-336 for SEQ ID No. 4, SEQ ID No. 15 and SEQ ID No. 16. (*A. salmonicida*). When used to determine the homology of identity of the amino acid sequences, it is preferred that the alignments as herein described use the mature sequence.

Therefore the most preferred regions for determining homology (identity) are amino acids 19-335 for SEQ ID No. 1 and 3 (*A. hydrophilia*) and amino acids 19-336 for SEQ ID No.s 4, 15 and 16. (*A. salmonicida*). SEQ ID 34 and 35 are mature protein sequences of a lipid acyl transferase from *A. hydrophilia* and *A. salmonicida* respectively.

A nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that may also be isolated from *Thermobifida*, preferably *T. fusca*, most preferably that encoded by SEQ ID No. 28.

A nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid

acyltransferase that may also be isolated from *Streptomyces*, preferable *S. avermitis*, most preferably that encoded by SEQ ID No. 32. Other possible enzymes for use in the present invention from *Streptomyces* include those encoded by SEQ ID No.s 5, 6, 9, 10, 11, 12, 13, 14, 31, and 33.

An enzyme for use in the invention may also be isolated from *Corynebacterium*, preferably *C. efficiens*, most preferably that encoded by SEQ ID No. 29.

Suitably, the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises any one of the amino acid sequences shown as SEQ ID No.s 37, 38, 40, 41, 43, 45, or 47 or an amino acid sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith, or encoded by any one of the nucleotide sequences shown as SEQ ID No.s 36, 39, 42, 44, 46, or 48 or a nucleotide sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith.

In one embodiment, the nucleic sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention is selected from the group consisting of:

a) a nucleic acid comprising a nucleotide sequence shown in SEQ ID No. 36;
b) a nucleic acid which is related to the nucleotide sequence of SEQ ID No. 36 by the degeneration of the genetic code; and
c) a nucleic acid comprising a nucleotide sequence which has at least 70% identity with the nucleotide sequence shown in SEQ ID No. 36.

In one embodiment, a nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that comprises an amino acid sequence as shown in SEQ ID No. 37 or an amino acid sequence which has at least 60% identity thereto.

In a further embodiment the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising any one of the amino acid sequences shown as SEQ ID No. 37, 38, 40, 41, 43, 45 or 47 or an amino acid

sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith, or encoded by any one of the nucleotide sequences shown as SEQ ID No. 39, 42, 44, 46 or 48 or a nucleotide sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith.

In a further embodiment the nucleotide sequence encoding a lipid acyltransferase enzyme for use any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising any one of amino sequences shown as SEQ ID No. 38, 40, 41, 45 or 47 or an amino acid sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith for the uses described herein.

In a further embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising any one of amino sequences shown as SEQ ID No. 38, 40, or 47 or an amino acid sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith for the uses described herein.

More preferably in one embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising the amino acid sequence shown as SEQ ID No. 47 or an amino acid sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith.

In another embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising the amino acid sequence shown as SEQ ID No. 43 or 44 or an amino acid sequence which has at least 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith.

In another embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase comprising the amino acid sequence shown as

SEQ ID No. 41 or an amino acid sequence which has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97% or 98% identity therewith.

In one embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention is selected from the group consisting of:

a) a nucleic acid comprising a nucleotide sequence shown in SEQ ID No. 36;
b) a nucleic acid which is related to the nucleotide sequence of SEQ ID No. 36 by the degeneration of the genetic code; and
c) a nucleic acid comprising a nucleotide sequence which has at least 70% identity with the nucleotide sequence shown in SEQ ID No. 36.

In one embodiment the lipid acyltransferase according to the present invention may be a lipid acyltransferase obtainable, preferably obtained, from the *Streptomyces* strains L130 or L131 deposited by Danisco A/S of Langebrogade 1, DK-1001 Copenhagen K, Denmark under the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the purposes of Patent Procedure at the National Collection of Industrial, Marine and Food Bacteria (NCIMB) 23 St. Machar Street, Aberdeen Scotland, GB on 25 June 2004 under accession numbers NCIMB 41226 and NCIMB 41227, respectively.

Suitable nucleotide sequences encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a polynucleotide encoding a lipid acyltransferase (SEQ ID No. 16); or may encode an amino acid sequence of a lipid acyltransferase (SEQ ID No. 17).

A suitable nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode an amino acid sequence which may be identified by alignment to the L131 (SEQ ID No. 37) sequence using Align X, the Clustal W pairwise alignment algorithm of VectorNTI using default settings.

An alignment of the L131 and homologues from *S. avermitilis* and *T. fusca* illustrates that the conservation of the GDSx motif (GDSY in L131 and *S. avermitilis* and *T. fusca*), the GANDY box, which is either GGNDA or GGNDL, and the HPT block

(considered to be the conserved catalytic histidine). These three conserved blocks are highlighted in Figure 42.

When aligned to either the pfam Pfam00657 consensus sequence (as described in WO04/064987) and/ or the L131 sequence herein disclosed (SEQ ID No 37) it is possible to identify three conserved regions, the GDSx block, the GANDY block and the HTP block (see WO04/064987 for further details).

When aligned to either the pfam Pfam00657 consensus sequence (as described in WO04/064987) and/ or the L131 sequence herein disclosed (SEQ ID No 37)

i) The nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that, has a GDSx motif, more preferably a GDSx motif selected from GDSL or GDSY motif.

and/or

ii) The nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that, has a GANDY block, more preferably a GANDY block comprising amino GGNDx, more preferably GGNDA or GGNDL.

and/or

iii) The nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyltransferase that has preferably an HTP block.

and preferably

iv) nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that has preferably a GDSx or GDSY motif, and a GANDY block comprising amino GGNDx, preferably GGNDA or GGNDL, and a HTP block (conserved histidine).

Variant lipid acyl transferase

In a preferred embodiment the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that is a variant lipid acyl transferase.

Variants which have an increased activity on phospholipids, such as increased hydrolytic activity and/ or increased transferase activity, preferably increased transferase activity on phospholipids may be used.

Preferably the variant lipid acyltransferase is prepared by one or more amino acid modifications of the lipid acyl transferases as defined hereinabove.

Suitably, when the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that may be a variant lipid acyltransferase, in which case the enzyme may be characterised in that the enzyme comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S, and wherein the variant enzyme comprises one or more amino acid modifications compared with a parent sequence at any one or more of the amino acid residues defined in set 2 or set 4 or set 6 or set 7 (as defined WO2005/066347 and hereinbelow).

For instance the variant lipid acyltransferase may be characterised in that the enzyme comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S, and wherein the variant enzyme comprises one or more amino acid modifications compared with a parent sequence at any one or more of the amino acid residues detailed in set 2 or set 4 or set 6 or set 7 (as defined in WO2005/066347 and hereinbelow) identified by said parent sequence being structurally aligned with the structural model of P10480 defined herein, which is preferably obtained by structural alignment of P10480 crystal structure coordinates with 1IVN.PDB and/or 1DEO.PDB as defined WO2005/066347 and hereinbelow.

In a further embodiment a nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a variant lipid acyltransferase that may be characterised in that the enzyme comprises the amino acid sequence motif GDSX, wherein X is one or more of the following amino acid residues L, A, V, I, F, Y, H, Q, T, N, M or S, and wherein the variant enzyme comprises one or more amino acid modifications compared with a parent sequence at any one or more of the amino acid residues taught in set 2 identified when said parent sequence is aligned to the pfam consensus sequence (SEQ ID No. 2 –Figure 3) and modified according to a structural model of P10480 to ensure best fit overlap as defined WO2005/066347 and hereinbelow.

Suitably the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a variant lipid acyltransferase enzyme that may comprise an amino acid sequence, which amino acid sequence is shown as SEQ ID No. 34, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 19, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, , SEQ ID No. 32, or SEQ ID No. 33 except for one or more amino acid modifications at any one or more of the amino acid residues defined in set 2 or set 4 or set 6 or set 7 (as defined WO2005/066347 and hereinbelow) identified by sequence alignment with SEQ ID No. 34.

Alternatively the nucleotide sequence encoding a lipid acyltransferase may encode a variant lipid acyltransferase enzyme comprising an amino acid sequence, which amino acid sequence is shown as SEQ ID No. 34, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 19, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, , SEQ ID No. 32, or SEQ ID No. 33 except for one or more amino acid modifications at any one or more of the amino acid residues defined in set 2 or set 4 or set 6 or set 7 as defined WO2005/066347 and hereinbelow, identified by said parent sequence being structurally aligned with the structural model of P10480 defined herein, which is preferably obtained by structural alignment of P10480 crystal

structure coordinates with 1IVN.PDB and/or 1DEO.PDB as taught within WO2005/066347 and hereinbelow.

Alternatively, the nucleotide sequence encoding a lipid acyltransferase may encode a variant lipid acyltransferase enzyme comprising an amino acid sequence, which amino acid sequence is shown as SEQ ID No. 34, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 19, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, , SEQ ID No. 32, or SEQ ID No. 33 except for one or more amino acid modifications at any one or more of the amino acid residues taught in set 2 identified when said parent sequence is aligned to the pfam consensus sequence (SEQ ID No. 2) and modified according to a structural model of P10480 to ensure best fit overlap as taught within WO2005/066347 and hereinbelow.

Preferably, the parent enzyme is an enzyme which comprises, or is homologous to, the amino acid sequence shown as SEQ ID No. 34 and/or SEQ ID No. 15 and/or SEQ ID No. 35.

Preferably, the nucleotide sequence encoding a lipid acyltransferase may encode a variant enzyme which comprises an amino acid sequence, which amino acid sequence is shown as SEQ ID No. 34 or SEQ ID No. 35 except for one or more amino acid modifications at any one or more of the amino acid residues defined in set 2 or set 4 or set 6 or set 7 as defined in WO2005/066347 and hereinbelow.

DEFINITION OF SETS

Amino acid set 1:

Amino acid set 1 (note that these are amino acids in 1IVN – Figure 53 and Figure 54) Gly8, Asp9, Ser10, Leu11, Ser12, Tyr15, Gly44, Asp45, Thr46, Glu69, Leu70, Gly71, Gly72, Asn73, Asp74, Gly75, Leu76, Gln106, Ile107, Arg108, Leu109, Pro110, Tyr113, Phe121, Phe139, Phe140, Met141, Tyr145, Met151, Asp154, His157, Gly155, Ile156, Pro158

The highly conserved motifs, such as GDSx and catalytic residues, were deselected from set 1 (residues underlined). For the avoidance of doubt, set 1 defines the amino acid residues within 10Å of the central carbon atom of a glycerol in the active site of the 1IVN model.

Amino acid set 2:

Amino acid set 2 (note that the numbering of the amino acids refers to the amino acids in the P10480 mature sequence)

Leu17, Lys22, Met23, Gly40, Asn80, Pro81, Lys82, Asn87, Asn88, Trp111, Val112, Ala114, Tyr117, Leu118, Pro156, Gly159, Gln160, Asn161, Pro162, Ser163, Ala164, Arg165, Ser166, Gln167, Lys168, Val169, Val170, Glu171, Ala172, Tyr179, His180, Asn181, Met209, Leu210, Arg211, Asn215, Lys284, Met285, Gln289 and Val290.

Table of selected residues in Set 1 compared with Set 2:

IVN model			P10480 Mature sequence Residue Number
IVN	A.hyd homologue		
	PFAM	Structure	
Gly8	Gly32		
Asp9	Asp33		
Ser10	Ser34		
Leu11	Leu35		Leu17
Ser12	Ser36		Ser18
			Lys22
			Met23
Tyr15	Gly58		Gly40
Gly44	Asn98		Asn80
Asp45	Pro99		Pro81
Thr46	Lys100		Lys82
			Asn87
			Asn88
Glu69	Trp129		Trp111
Leu70	Val130		Val112

Gly71	Gly131		
Gly72	Ala132		Ala114
Asn73	Asn133		
Asp74	Asp134		
Gly75	Tyr135		Tyr117
Leu76	Leu136		Leu118
Gln106		Pro174	Pro156
Ile107		Gly177	Gly159
Arg108		Gln178	Gln160
Leu109		Asn179	Asn161
Pro110		180 to 190	Pro162
Tyr113			Ser163
			Ala164
			Arg165
			Ser166
			Gln167
			Lys168
			Val169
			Val170
			Glu171
			Ala172
Phe121	His198	Tyr197	Tyr179
		His198	His180
		Asn199	Asn181
Phe139	Met227		Met209
Phe140	Leu228		Leu210
Met141	Arg229		Arg211
Tyr145	Asn233		Asn215
			Lys284
Met151	Met303		Met285
Asp154	Asp306		
Gly155	Gln307		Gln289
Ile156	Val308		Val290
His157	His309		

Pro158	Pro310		

Amino acid set 3:

Amino acid set 3 is identical to set 2 but refers to the *Aeromonas salmonicida* (SEQ ID No. 4) coding sequence, i.e. the amino acid residue numbers are 18 higher in set 3 as this reflects the difference between the amino acid numbering in the mature protein (SEQ ID No. 34) compared with the protein including a signal sequence (SEQ ID No. 25).

The mature proteins of *Aeromonas salmonicida* GDSX (SEQ ID No. 4) and *Aeromonas hydrophila* GDSX (SEQ ID No. 34) differ in five amino acids. These are Thr3Ser, Gln182Lys, Glu309Ala, Ser310Asn, and Gly318-, where the *salmonicida* residue is listed first and the *hydrophila* residue is listed last. The *hydrophila* protein is only 317 amino acids long and lacks a residue in position 318. The *Aeromonas salmonicida* GDSX has considerably high activity on polar lipids such as galactolipid substrates than the *Aeromonas hydrophila* protein. Site scanning was performed on all five amino acid positions.

Amino acid set 4:

Amino acid set 4 is S3, Q182, E309, S310, and -318.

Amino acid set 5:

F13S, D15N, S18G, S18V, Y30F, D116N, D116E, D157 N, Y226F, D228N Y230F.

Amino acid set 6:

Amino acid set 6 is Ser3, Leu17, Lys22, Met23, Gly40, Asn80, Pro81, Lys82, Asn 87, Asn88, Trp111, Val112, Ala114, Tyr117, Leu118, Pro156, Gly159, Gln160, Asn161, Pro162, Ser163, Ala164, Arg165, Ser166, Gln167, Lys168, Val169, Val170, Glu171, Ala172, Tyr179, His180, Asn181, Gln182, Met209, Leu210, Arg211, Asn215, Lys284, Met285, Gln289, Val290, Glu309, Ser310, -318.

The numbering of the amino acids in set 6 refers to the amino acids residues in P10480 (SEQ ID No. 25) – corresponding amino acids in other sequence backbones can be determined by homology alignment and/or structural alignment to P10480 and/or 1IVN.

Amino acid set 7:

Amino acid set 7 is Ser3, Leu17, Lys22, Met23, Gly40, Asn80, Pro81, Lys82, Asn 87, Asn88, Trp111, Val112, Ala114, Tyr117, Leu118, Pro156, Gly159, Gln160, Asn161, Pro162, Ser163, Ala164, Arg165, Ser166, Gln167, Lys168, Val169, Val170, Glu171, Ala172, Tyr179, His180, Asn181, Gln182, Met209, Leu210, Arg211, Asn215, Lys284, Met285, Gln289, Val290, Glu309, Ser310, -318, Y30X (where X is selected from A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W), Y226X (where X is selected from A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W), Y230X (where X is selected from A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W), S18X (where X is selected from A, C, D, E, F, H, I, K, L, M, N, P, Q, R, T, W or Y), D157X (where X is selected from A, C, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W or Y).

The numbering of the amino acids in set 7 refers to the amino acids residues in P10480 (SEQ ID No. 25) – corresponding amino acids in other sequence backbones can be determined by homology alignment and/or structural alignment to P10480 and/or 1IVN).

Suitably, the variant enzyme comprises one or more of the following amino acid modifications compared with the parent enzyme:
S3E, A, G, K, M, Y, R, P, N, T or G
E309Q, R or A, preferably Q or R
-318Y, H, S or Y, preferably Y.

Preferably, X of the GDSX motif is L. Thus, preferably the parent enzyme comprises the amino acid motif GDSL.

Suitably, said first parent lipid acyltransferase may comprise any one of the following amino acid sequences: SEQ ID No. 34, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 19, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, , SEQ ID No. 32 or SEQ ID No. 33.

Suitably, said second related lipid acyltransferase may comprise any one of the following amino acid sequences: SEQ ID No. 3, SEQ ID No. 34, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 19, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 14, SEQ ID No. 1, SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, , SEQ ID No. 32 or SEQ ID No. 33.

The variant enzyme must comprise at least one amino acid modification compared with the parent enzyme. In some embodiments, the variant enzyme may comprise at least 2, preferably at least 3, preferably at least 4, preferably at least 5, preferably at least 6, preferably at least 7, preferably at least 8, preferably at least 9, preferably at least 10 amino acid modifications compared with the parent enzyme.

When referring to specific amino acid residues herein the numbering is that obtained from alignment of the variant sequence with the reference sequence shown as SEQ ID No. 34 or SEQ ID No. 35.

In one aspect preferably the variant enzyme comprises one or more of the following amino acid substitutions:

S3A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W, or Y; and/or
L17A, C, D, E, F, G, H, I, K, M, N, P, Q, R, S, T, V, W, or Y; and/or
S18A, C, D, E, F, H, I, K, L, M, N, P, Q, R, T, W, or Y; and/or
K22A, C, D, E, F, G, H, I, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
M23A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W, or Y; and/or
Y30A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W; and/or
G40A, C, D, E, F, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
N80A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
P81A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, or Y; and/or
K82A, C, D, E, F, G, H, I, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
N87A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
N88A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
W111A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W or Y; and/or
V112A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, W, or Y; and/or
A114C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
Y117A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W; and/or
L118A, C, D, E, F, G, H, I, K, M, N, P, Q, R, S, T, V, W, or Y; and/or
P156A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, or Y; and/or
D157A, C, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
G159A, C, D, E, F, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
Q160A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W, or Y; and/or
N161A, C, D, E, F, G, H, I, K, L, M P, Q, R, S, T, V, W, or Y; and/or
P162A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, or Y; and/or
S163A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W, or Y; and/or
A164C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
R165A, C, D, E, F, G, H, I, K, L, M, N, P, Q, S, T, V, W, or Y; and/or
S166A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W, or Y; and/or
Q167A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W, or Y; and/or

K168A, C, D, E, F, G, H, I, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
V169A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, W, or Y; and/or
V170A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, W, or Y; and/or
E171A, C, D, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
A172C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
Y179A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W; and/or
H180A, C, D, E, F, G, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
N181A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, or Y; and/or
Q182A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W, or Y, preferably K; and/or
M209A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W, or Y; and/or
L210 A, C, D, E, F, G, H, I, K, M, N, P, Q, R, S, T, V, W, or Y; and/or
R211 A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
N215 A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
Y226A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, or W; and/or
Y230A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V or W; and/or
K284A, C, D, E, F, G, H, I, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
M285A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W, or Y; and/or
Q289A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W, or Y; and/or
V290A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, W, or Y; and/or
E309A, C, D, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, or Y; and/or
S310A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W, or Y.

In addition or alternatively thereto there may be one or more C-terminal extensions. Preferably the additional C-terminal extension is comprised of one or more aliphatic amino acids, preferably a non-polar amino acid, more preferably of I, L, V or G. Thus, the present invention further provides for a variant enzyme comprising one or more of the following C-terminal extensions: 318I, 318L, 318V, 318G.

Preferred variant enzymes may have a decreased hydrolytic activity against a phospholipid, such as phosphatidylcholine (PC), may also have an increased transferase activity from a phospholipid.

Preferred variant enzymes may have an increased transferase activity from a phospholipid, such as phosphatidylcholine (PC), these may also have an increased hydrolytic activity against a phospholipid.

Modification of one or more of the following residues may result in a variant enzyme having an increased absolute transferase activity against phospholipid:

S3, D157, S310, E309, Y179, N215, K22, Q289, M23, H180, M209, L210, R211, P81, V112, N80, L82, N88; N87

Specific preferred modifications which may provide a variant enzyme having an improved transferase activity from a phospholipid may be selected from one or more of the following:

S3A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y; preferably N, E, K, R, A, P or M, most preferably S3A

D157A, C, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W or Y; preferably D157S, R, E, N, G, T, V, Q, K or C

S310A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y; preferably S310T

-318 E

E309A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y; preferably E309 R, E, L, R or A

Y179A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V or W; preferably Y179 D, T, E, R, N, V, K, Q or S, more preferably E, R, N, V, K or Q

N215A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W or Y; preferably N215 S, L, R or Y

K22A, C, D, E, F, G, H, I, L, M, N, P, Q, R, S, T, V, W or Y; preferably K22 E, R, C or A

Q289A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W or Y; preferably Q289 R, E, G, P or N

M23A, C, D, E, F, G, H, I, K, L N, P, Q, R, S, T, V, W or Y; preferably M23 K, Q, L, G, T or S

H180A, C, D, E, F, G, I, K, L, M, P, Q, R, S, T, V, W or Y; preferably H180 Q, R or K

M209 A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W or Y; preferably M209 Q, S, R, A, N, Y, E, V or L

L210A, C, D, E, F, G, H, I, K, M, N, P, Q, R, S, T, V, W or Y; preferably L210 R, A, V, S, T, I, W or M

R211A, C, D, E, F, G, H, I, K, L, M, N, P, Q, S, T, V, W or Y; preferably R211T

P81A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W or Y; preferably P81G

V112A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, W or Y; preferably V112C
N80A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W or Y; preferably N80 R, G, N, D, P, T, E, V, A or G
L82A, C, D, E, F, G, H, I, M, N, P, Q, R, S, T, V, W or Y; preferably L82N, S or E
N88A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W or Y; preferably N88C
N87A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W or Y; preferably N87M or G

Preferred modification of one or more of the following residues results in a variant enzyme having an increased absolute transferase activity against phospholipid:

S3 N, R, A, G
M23 K, Q, L, G, T, S
H180 R
L82 G
Y179 E, R, N, V, K or Q
E309 R, S, L or A

One preferred modification is N80D. This is particularly the case when using the reference sequence SEQ ID No. 35 as the backbone. Thus, the reference sequence may be SEQ ID No. 16. This modification may be in combination with one or more further modifications. Therefore in a preferred embodiment of the present invention the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid acyltransferase that comprises SEQ ID No. 35 or an amino acid sequence which has 75% or more, preferably 85% or more, more preferably 90% or more, even more preferably 95% or more, even more preferably 98% or more, or even more preferably 99% or more identity to SEQ ID No. 35.

As noted above, when referring to specific amino acid residues herein the numbering is that obtained from alignment of the variant sequence with the reference sequence shown as SEQ ID No. 34 or SEQ ID No. 35

Much by preference, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention may encode a lipid comprising the amino acid sequence shown as SEQ ID No. 16, or an

amino acid sequence which has 75% or more, preferably 85% or more, more preferably 90% or more, even more preferably 95% or more, even more preferably 98% or more, or even more preferably 99% or more identity to SEQ ID No. 16. This enzyme may be considered a variant enzyme.

For the purposes of the present invention, the degree of identity is based on the number of sequence elements which are the same. The degree of identity in accordance with the present invention for amino acid sequences may be suitably determined by means of computer programs known in the art, such as Vector NTI 10 (Invitrogen Corp.). For pairwise alignment the score used is preferably BLOSUM62 with Gap opening penalty of 10.0 and Gap extension penalty of 0.1.

Suitably, the degree of identity with regard to an amino acid sequence is determined over at least 20 contiguous amino acids, preferably over at least 30 contiguous amino acids, preferably over at least 40 contiguous amino acids, preferably over at least 50 contiguous amino acids, preferably over at least 60 contiguous amino acids.

Suitably, the degree of identity with regard to an amino acid sequence may be determined over the whole sequence.

Suitably, the nucleotide sequence encoding a lipid acyltransferase/ lipid acyl transferase enzyme according to the present invention may be obtainable, preferably obtained, from organisms from one or more of the following genera: *Aeromonas, Streptomyces, Saccharomyces, Lactococcus, Mycobacterium, Streptococcus, Lactobacillus, Desulfitobacterium, Bacillus, Campylobacter, Vibrionaceae, Xylella, Sulfolobus, Aspergillus, Schizosaccharomyces, Listeria, Neisseria, Mesorhizobium, Ralstonia, Xanthomonas, Candida, Thermobifida* and *Corynebacterium.*

Suitably, the nucleotide sequence encoding a lipid acyltransferase / lipid acyl transferase enzyme according to the present invention may be obtainable, preferably obtained, from one or more of the following organisms: *Aeromonas hydrophila, Aeromonas salmonicida, Streptomyces coelicolor, Streptomyces rimosus, Mycobacterium, Streptococcus pyogenes, Lactococcus lactis, Streptococcus pyogenes, Streptococcus thermophilus, Streptomyces thermosacchari, Streptomyces avermitilis Lactobacillus helveticus, Desulfitobacterium dehalogenans, Bacillus sp,*

Campylobacter jejuni, Vibrionaceae, Xylella fastidiosa, Sulfolobus solfataricus, Saccharomyces cerevisiae, Aspergillus terreus, Schizosaccharomyces pombe, Listeria innocua, Listeria monocytogenes, Neisseria meningitidis, Mesorhizobium loti, Ralstonia solanacearum, Xanthomonas campestris, Xanthomonas axonopodis , Candida parapsilosis Thermobifida fusca and *Corynebacterium efficiens.*

In one aspect, preferably the nucleotide sequence encoding a nucleotide sequence encoding a lipid acyltransferase for use in any one of the host cells, vectors, methods and uses of the present invention encodes a lipid acyl transferase enzyme according to the present invention is obtainable, preferably obtained or derived, from one or more of *Aeromonas spp., Aeromonas hydrophila* or *Aeromonas salmonicida*.

Enzymes which function as lipid acyltransferases in accordance with the present invention can be routinely identified using the assay taught in Example 12 of WO2004/064537. Using this assay, in which there is a very high water content – approximately 95%, lipid acyltransferases/lipid acyl transferase in accordance with the present invention are those which have at least 2% acyltransferase activity (relative transferase activity), preferably at least 5% relative transferase activity, preferably at least 10% relative transferase activity, preferably at least 15%, 20%, 25% 26%, 28%, 30%, 40% 50%, 60% or 75% relative transferase activity.

Phospholipases may act as acyl-transferase enzymes in low water environments. Therefore it is considered that in place of or in addition to the phospholipid acyltransferase enzyme a phospholipase enzyme may be used when process for the modification of the edible oil of fat takes place in a low water environment.

The term "high water" as used herein means any substrate or foodstuff with more than 3% water content, preferably more than 4%, more than 5%, more than 6%, more than 7%, more than 8%, more than 9%, more than 10%, more than 20%, more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80% or more than 90%.

The term "low water" as used herein means any substrate or foodstuff with less than 3% water content, preferably less than 2%, less than 1% or less than 0.5%, less than 0.3%, less than 0.2, less than 0.1, less than 0.05, or less than 0.01%

For avoidance of doubt milk is a high water environment where as butterfat is a low water environment.

Suitable phospholipases for use in the invention include phospholipase A1, phospholipase A2, or phospholipase B. Phospholipase A1, phospholipase A2, or phospholipase B may also be used in co-ordination with the lipid acyl transferase activity. Phospholipase C and /or D may also be used in co-ordination with the lipid acyl transferase activity/phospholipase A1, A2 and/or B activity in analogy with WO2005/089562. Preferred phospholipases may include phospholipase A2, such as Lecitase*TM* or the *Fusarium venenatum* and *Tuber albidum* phospholipase disclosed in WO2004/97012 (Novozymes/Chr. Hansen). A *Fusarium venenatum* phospholipase is sold by Novozymes as MAX YIELD™.

ISOLATED

In one aspect, the method of the present invention comprises the additional step of recovering/isolating the lipid acyltransferase. Thus, the lipid acyltransferase produced may be in an isolated form.

In another aspect, the nucleotide sequence encoding a lipid acyltransferase for use in the present invention may be in an isolated form.

The term "isolated" means that the sequence or protein is at least substantially free from at least one other component with which the sequence or protein is naturally associated in nature and as found in nature.

PURIFIED

In one aspect, the method of the present invention comprises the additional step of purifying the lipid acyltransferase.

In another aspect, the nucleotide sequence encoding a lipid acyltransferase for use in the present invention may be in a purified form.

The term "purified" means that the sequence is in a relatively pure state – e.g. at least about 51% pure, or at least about 75%, or at least about 80%, or at least about 90% pure, or at least about 95% pure or at least about 98% pure.

CLONING A NUCLEOTIDE SEQUENCE ENCODING A POLYPEPTIDE ACCORDING TO THE PRESENT INVENTION

A nucleotide sequence encoding either a polypeptide which has the specific properties as defined herein or a polypeptide which is suitable for modification may be isolated from any cell or organism producing said polypeptide. Various methods are well known within the art for the isolation of nucleotide sequences.

For example, a genomic DNA and/or cDNA library may be constructed using chromosomal DNA or messenger RNA from the organism producing the polypeptide. If the amino acid sequence of the polypeptide is known, labeled oligonucleotide probes may be synthesised and used to identify polypeptide-encoding clones from the genomic library prepared from the organism. Alternatively, a labelled oligonucleotide probe containing sequences homologous to another known polypeptide gene could be used to identify polypeptide-encoding clones. In the latter case, hybridisation and washing conditions of lower stringency are used.

Alternatively, polypeptide-encoding clones could be identified by inserting fragments of genomic DNA into an expression vector, such as a plasmid, transforming enzyme-negative bacteria with the resulting genomic DNA library, and then plating the transformed bacteria onto agar containing an enzyme inhibited by the polypeptide, thereby allowing clones expressing the polypeptide to be identified.

In a yet further alternative, the nucleotide sequence encoding the polypeptide may be prepared synthetically by established standard methods, e.g. the phosphoroamidite method described by Beucage S.L. *et al* (1981) Tetrahedron Letters 22, p 1859-1869, or the method described by Matthes *et al* (1984) EMBO J. 3, p 801-805. In the phosphoroamidite method, oligonucleotides are synthesised, e.g. in an automatic DNA synthesiser, purified, annealed, ligated and cloned in appropriate vectors.

The nucleotide sequence may be of mixed genomic and synthetic origin, mixed synthetic and cDNA origin, or mixed genomic and cDNA origin, prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate) in accordance with standard techniques. Each ligated fragment corresponds to various parts of the entire nucleotide sequence. The DNA sequence may also be prepared by polymerase chain reaction (PCR) using specific primers, for instance as described in US 4,683,202 or in Saiki R K *et al* (Science (1988) 239, pp 487-491).

NUCLEOTIDE SEQUENCES

The present invention also encompasses nucleotide sequences encoding polypeptides having the specific properties as defined herein. The term "nucleotide sequence" as used herein refers to an oligonucleotide sequence or polynucleotide sequence, and variant, homologues, fragments and derivatives thereof (such as portions thereof). The nucleotide sequence may be of genomic or synthetic or recombinant origin, which may be double-stranded or single-stranded whether representing the sense or antisense strand.

The term "nucleotide sequence" in relation to the present invention includes genomic DNA, cDNA, synthetic DNA, and RNA. Preferably it means DNA, more preferably cDNA for the coding sequence.

In a preferred embodiment, the nucleotide sequence *per se* encoding a polypeptide having the specific properties as defined herein does not cover the native nucleotide sequence in its natural environment when it is linked to its naturally associated sequence(s) that is/are also in its/their natural environment. For ease of reference, we shall call this preferred embodiment the "non-native nucleotide sequence". In this regard, the term "native nucleotide sequence" means an entire nucleotide sequence that is in its native environment and when operatively linked to an entire promoter with which it is naturally associated, which promoter is also in its native environment. Thus, the polypeptide of the present invention can be expressed by a nucleotide sequence in its native organism but wherein the nucleotide sequence is not under the control of the promoter with which it is naturally associated within that organism.

Preferably the polypeptide is not a native polypeptide. In this regard, the term "native polypeptide" means an entire polypeptide that is in its native environment and when it has been expressed by its native nucleotide sequence.

Typically, the nucleotide sequence encoding polypeptides having the specific properties as defined herein is prepared using recombinant DNA techniques (i.e. recombinant DNA). However, in an alternative embodiment of the invention, the nucleotide sequence could be synthesised, in whole or in part, using chemical methods well known in the art (see Caruthers MH *et al* (1980) Nuc Acids Res Symp Ser 215-23 and Horn T *et al* (1980) Nuc Acids Res Symp Ser 225-232).

MOLECULAR EVOLUTION

Once an enzyme-encoding nucleotide sequence has been isolated, or a putative enzyme-encoding nucleotide sequence has been identified, it may be desirable to modify the selected nucleotide sequence, for example it may be desirable to mutate the sequence in order to prepare an enzyme in accordance with the present invention.

Mutations may be introduced using synthetic oligonucleotides. These oligonucleotides contain nucleotide sequences flanking the desired mutation sites.

A suitable method is disclosed in Morinaga *et al* (Biotechnology (1984) 2, p646-649). Another method of introducing mutations into enzyme-encoding nucleotide sequences is described in Nelson and Long (Analytical Biochemistry (1989), 180, p 147-151).

Instead of site directed mutagenesis, such as described above, one can introduce mutations randomly for instance using a commercial kit such as the GeneMorph PCR mutagenesis kit from Stratagene, or the Diversify PCR random mutagenesis kit from Clontech. EP 0 583 265 refers to methods of optimising PCR based mutagenesis, which can also be combined with the use of mutagenic DNA analogues such as those described in EP 0 866 796. Error prone PCR technologies are suitable for the production of variants of lipid acyl transferases with preferred characteristics. WO0206457 refers to molecular evolution of lipases.

A third method to obtain novel sequences is to fragment non-identical nucleotide sequences, either by using any number of restriction enzymes or an enzyme such as Dnase I, and reassembling full nucleotide sequences coding for functional proteins. Alternatively one can use one or multiple non-identical nucleotide sequences and introduce mutations during the reassembly of the full nucleotide sequence. DNA shuffling and family shuffling technologies are suitable for the production of variants of lipid acyl transferases with preferred characteristics. Suitable methods for performing 'shuffling' can be found in EP0 752 008, EP1 138 763, EP1 103 606. Shuffling can also be combined with other forms of DNA mutagenesis as described in US 6,180,406 and WO 01/34835.

Thus, it is possible to produce numerous site directed or random mutations into a nucleotide sequence, either *in vivo* or *in vitro*, and to subsequently screen for improved functionality of the encoded polypeptide by various means. Using in silico and exo mediated recombination methods (see WO 00/58517, US 6,344,328, US 6,361,974), for example, molecular evolution can be performed where the variant produced retains very low homology to known enzymes or proteins. Such variants thereby obtained may have significant structural analogy to known transferase enzymes, but have very low amino acid sequence homology.

As a non-limiting example, In addition, mutations or natural variants of a polynucleotide sequence can be recombined with either the wild type or other mutations or natural variants to produce new variants. Such new variants can also be screened for improved functionality of the encoded polypeptide.

The application of the above-mentioned and similar molecular evolution methods allows the identification and selection of variants of the enzymes of the present invention which have preferred characteristics without any prior knowledge of protein structure or function, and allows the production of non-predictable but beneficial mutations or variants. There are numerous examples of the application of molecular evolution in the art for the optimisation or alteration of enzyme activity, such examples include, but are not limited to one or more of the following: optimised expression and/or activity in a host cell or in vitro, increased enzymatic activity, altered substrate and/or product specificity, increased or decreased enzymatic or structural stability,

altered enzymatic activity/specificity in preferred environmental conditions, e.g. temperature, pH, substrate

As will be apparent to a person skilled in the art, using molecular evolution tools an enzyme may be altered to improve the functionality of the enzyme.

Suitably, the nucleotide sequence encoding a lipid acyltransferase used in the invention may encode a variant lipid acyltransferase, i.e. the lipid acyltransferase may contain at least one amino acid substitution, deletion or addition, when compared to a parental enzyme. Variant enzymes retain at least 1%, 2%, 3%, 5%, 10%, 15%, 20%, 30%, 40%, 50 %, 60%, 70%, 80%, 90%, 95%, 97%, 99% homology with the parent enzyme. Suitable parent enzymes may include any enzyme with esterase or lipase activity. Preferably, the parent enzyme aligns to the pfam00657 consensus sequence.

In a preferable embodiment a variant lipid acyltransferase enzyme retains or incorporates at least one or more of the pfam00657 consensus sequence amino acid residues found in the GDSx, GANDY and HPT blocks.

Enzymes, such as lipases with no or low lipid acyltransferase activity in an aqueous environment may be mutated using molecular evolution tools to introduce or enhance the transferase activity, thereby producing a lipid acyltransferase enzyme with significant transferase activity suitable for use in the compositions and methods of the present invention.

Suitably, the nucleotide sequence encoding a lipid acyltransferase for use in any one of the vectors, host cells, methods and/or uses of the present invention may encode a lipid acyltransferase that may be a variant with enhanced enzyme activity on polar lipids, preferably phospholipids and/or glycolipids when compared to the parent enzyme. Preferably, such variants also have low or no activity on lyso polar lipids. The enhanced activity on polar lipids, phospholipids and/or glycolipids may be the result of hydrolysis and/or transferase activity or a combination of both.

Variant lipid acyltransferases may have decreased activity on triglycerides, and/or monoglycerides and/or diglycerides compared with the parent enzyme.

Suitably the variant enzyme may have no activity on triglycerides and/or monoglycerides and/or diglycerides.

Alternatively, the variant enzyme may have increased activity on triglycerides, and/or may also have increased activity on one or more of the following, polar lipids, phospholipids, lecithin, phosphatidylcholine, glycolipids, digalactosyl monoglyceride, monogalactosyl monoglyceride.

Variants of lipid acyltransferases are known, and one or more of such variants may be suitable for use in the methods and uses according to the present invention and/or in the enzyme compositions according to the present invention. By way of example only, variants of lipid acyltransferases are described in the following references may be used in accordance with the present invention: Hilton & Buckley J Biol. Chem. 1991 Jan 15: 266 (2): 997-1000; Robertson *et al* J. Biol. Chem. 1994 Jan 21; 269(3):2146-50; Brumlik *et al* J. Bacteriol 1996 Apr; 178 (7): 2060-4; Peelman *et al* Protein Sci. 1998 Mar; 7(3):587-99.

AMINO ACID SEQUENCES

The present invention also encompasses amino acid sequences encoded by a nucleotide sequence which encodes a lipid acyltransferase for use in any one of the vectors, host cells, methods and/or uses of the present invention.

As used herein, the term "amino acid sequence" is synonymous with the term "polypeptide" and/or the term "protein". In some instances, the term "amino acid sequence" is synonymous with the term "peptide".

The amino acid sequence may be prepared/isolated from a suitable source, or it may be made synthetically or it may be prepared by use of recombinant DNA techniques.

Suitably, the amino acid sequences may be obtained from the isolated polypeptides taught herein by standard techniques.

One suitable method for determining amino acid sequences from isolated polypeptides is as follows:

Purified polypeptide may be freeze-dried and 100 µg of the freeze-dried material may be dissolved in 50 µl of a mixture of 8 M urea and 0.4 M ammonium hydrogen carbonate, pH 8.4. The dissolved protein may be denatured and reduced for 15 minutes at 50°C following overlay with nitrogen and addition of 5 µl of 45 mM dithiothreitol. After cooling to room temperature, 5 µl of 100 mM iodoacetamide may be added for the cysteine residues to be derivatized for 15 minutes at room temperature in the dark under nitrogen.

135 µl of water and 5 µg of endoproteinase Lys-C in 5 µl of water may be added to the above reaction mixture and the digestion may be carried out at 37°C under nitrogen for 24 hours.

The resulting peptides may be separated by reverse phase HPLC on a VYDAC C18 column (0.46x15cm;10µm; The Separation Group, California, USA) using solvent A: 0.1% TFA in water and solvent B: 0.1% TFA in acetonitrile. Selected peptides may be re-chromatographed on a Develosil C18 column using the same solvent system, prior to N-terminal sequencing. Sequencing may be done using an Applied Biosystems 476A sequencer using pulsed liquid fast cycles according to the manufacturer's instructions (Applied Biosystems, California, USA).

SEQUENCE IDENTITY OR SEQUENCE HOMOLOGY

Here, the term "homologue" means an entity having a certain homology with the subject amino acid sequences and the subject nucleotide sequences. Here, the term "homology" can be equated with "identity".

The homologous amino acid sequence and/or nucleotide sequence should provide and/or encode a polypeptide which retains the functional activity and/or enhances the activity of the enzyme.

In the present context, a homologous sequence is taken to include an amino acid sequence which may be at least 75, 85 or 90% identical, preferably at least 95 or 98%

identical to the subject sequence. Typically, the homologues will comprise the same active sites etc. as the subject amino acid sequence. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

In the present context, a homologous sequence is taken to include a nucleotide sequence which may be at least 75, 85 or 90% identical, preferably at least 95 or 98% identical to a nucleotide sequence encoding a polypeptide of the present invention (the subject sequence). Typically, the homologues will comprise the same sequences that code for the active sites etc. as the subject sequence. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the Vector NTI (Invitrogen Corp.). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al* 1999 Short Protocols in Molecular Biology, 4th Ed – Chapter 18), and FASTA (Altschul *et al* 1990 J. Mol. Biol. 403-410). Both BLAST and FASTA are available for offline and online searching (see Ausubel *et al* 1999, pages 7-58 to 7-60). However, for some applications, it is preferred to use the Vector NTI program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequence (see FEMS Microbiol Lett 1999 174(2): 247-50; FEMS Microbiol Lett 1999 177(1): 187-8 and tatiana@ncbi.nlm.nih.gov).

Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. Vector NTI programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). For some applications, it is preferred to use the default values for the Vector NTI package.

Alternatively, percentage homologies may be calculated using the multiple alignment feature in Vector NTI (Invitrogen Corp.), based on an algorithm, analogous to CLUSTAL (Higgins DG & Sharp PM (1988), *Gene* 73(1), 237-244).

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

Should Gap Penalties be used when determining sequence identity, then preferably the following parameters are used for pairwise alignment:

FOR BLAST	
GAP OPEN	0
GAP EXTENSION	0

FOR CLUSTAL	DNA	PROTEIN	
WORD SIZE	2	1	K triple
GAP PENALTY	15	10	
GAP EXTENSION	6.66	0.1	

In one embodiment, preferably the sequence identity for the nucleotide sequences is determined using CLUSTAL with the gap penalty and gap extension set as defined above.

Suitably, the degree of identity with regard to a nucleotide sequence is determined over at least 20 contiguous nucleotides, preferably over at least 30 contiguous nucleotides, preferably over at least 40 contiguous nucleotides, preferably over at least 50 contiguous nucleotides, preferably over at least 60 contiguous nucleotides, preferably over at least 100 contiguous nucleotides.

Suitably, the degree of identity with regard to a nucleotide sequence may be determined over the whole sequence.

In one embodiment the degree of amino acid sequence identity in accordance with the present invention may be suitably determined by means of computer programs known in the art, such as Vector NTI 10 (Invitrogen Corp.). For pairwise alignment the matrix used is preferably BLOSUM62 with Gap opening penalty of 10.0 and Gap extension penalty of 0.1.

Suitably, the degree of identity with regard to an amino acid sequence is determined over at least 20 contiguous amino acids, preferably over at least 30 contiguous amino acids, preferably over at least 40 contiguous amino acids, preferably over at least 50 contiguous amino acids, preferably over at least 60 contiguous amino acids.

Suitably, the degree of identity with regard to an amino acid sequence may be determined over the whole sequence.

The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar – uncharged	C S T M
		N Q
	Polar – charged	D E
		K R
AROMATIC		H F W Y

The present invention also encompasses homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue, with an alternative residue) that may occur i.e. like-for-like substitution such as basic for basic, acidic for acidic, polar for polar etc. Non-homologous substitution may also occur i.e. from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine.

Replacements may also be made by unnatural amino acids.

Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β-alanine residues. A further form of variation, involves the presence of one or more amino acid residues in peptoid form, will be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α-carbon substituent group is on the residue's nitrogen atom rather than the α-carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon RJ et al., PNAS (1992) 89(20), 9367-9371 and Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134.

Nucleotide sequences for use in the present invention or encoding a polypeptide having the specific properties defined herein may include within them synthetic or

modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones and/or the addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the nucleotide sequences described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of nucleotide sequences.

The present invention also encompasses the use of nucleotide sequences that are complementary to the sequences discussed herein, or any derivative, fragment or derivative thereof. If the sequence is complementary to a fragment thereof then that sequence can be used as a probe to identify similar coding sequences in other organisms etc.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other variants of the sequences described herein may be obtained for example by probing DNA libraries made from a range of individuals, for example individuals from different populations. In addition, other viral/bacterial, or cellular homologues particularly cellular homologues found in mammalian cells (e.g. rat, mouse, bovine and primate cells), may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the sequences shown in the sequence listing herein. Such sequences may be obtained by probing cDNA libraries made from or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of any one of the sequences in the attached sequence listings under conditions of medium to high stringency. Similar considerations apply to obtaining species homologues and allelic variants of the polypeptide or nucleotide sequences of the invention.

Variants and strain/species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding conserved amino acid sequences within the sequences of the present invention. Conserved sequences can be predicted, for example, by aligning the amino acid sequences from several variants/homologues. Sequence alignments can be

performed using computer software known in the art. For example the GCG Wisconsin PileUp program is widely used.

The primers used in degenerate PCR will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of characterised sequences. This may be useful where for example silent codon sequence changes are required to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction polypeptide recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

Polynucleotides (nucleotide sequences) of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Polynucleotides such as DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a stepwise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the lipid targeting sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain

reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

HYBRIDISATION

The present invention also encompasses sequences that are complementary to the sequences of the present invention or sequences that are capable of hybridising either to the sequences of the present invention or to sequences that are complementary thereto.

The term "hybridisation" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as the process of amplification as carried out in polymerase chain reaction (PCR) technologies.

The present invention also encompasses the use of nucleotide sequences that are capable of hybridising to the sequences that are complementary to the subject sequences discussed herein, or any derivative, fragment or derivative thereof.

The present invention also encompasses sequences that are complementary to sequences that are capable of hybridising to the nucleotide sequences discussed herein.

Hybridisation conditions are based on the melting temperature (Tm) of the nucleotide binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol. 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about Tm-5°C (5°C below the Tm of the probe); high stringency at about 5°C to 10°C below Tm; intermediate stringency at about 10°C to 20°C below Tm; and low stringency at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridisation

can be used to identify or detect identical nucleotide sequences while an intermediate (or low) stringency hybridisation can be used to identify or detect similar or related polynucleotide sequences.

Preferably, the present invention encompasses sequences that are complementary to sequences that are capable of hybridising under high stringency conditions or intermediate stringency conditions to nucleotide sequences encoding polypeptides having the specific properties as defined herein.

More preferably, the present invention encompasses sequences that are complementary to sequences that are capable of hybridising under high stringent conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na-citrate pH 7.0}) to nucleotide sequences encoding polypeptides having the specific properties as defined herein.

The present invention also relates to nucleotide sequences that can hybridise to the nucleotide sequences discussed herein (including complementary sequences of those discussed herein).

The present invention also relates to nucleotide sequences that are complementary to sequences that can hybridise to the nucleotide sequences discussed herein (including complementary sequences of those discussed herein).

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridising to the nucleotide sequences discussed herein under conditions of intermediate to maximal stringency.

In a preferred aspect, the present invention covers nucleotide sequences that can hybridise to the nucleotide sequences discussed herein, or the complement thereof, under stringent conditions (e.g. 50°C and 0.2xSSC).

In a more preferred aspect, the present invention covers nucleotide sequences that can hybridise to the nucleotide sequences discussed herein, or the complement thereof, under high stringent conditions (e.g. 65°C and 0.1xSSC).

EXPRESSION OF POLYPEPTIDES

A nucleotide sequence for use in the present invention or for encoding a polypeptide having the specific properties as defined herein can be incorporated into a recombinant replicable vector. The vector may be used to replicate and express the nucleotide sequence, in polypeptide form, in and/or from a compatible host cell. Expression may be controlled using control sequences which include promoters/enhancers and other expression regulation signals. Prokaryotic promoters and promoters functional in eukaryotic cells may be used. Tissue specific or stimuli specific promoters may be used. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

The polypeptide produced by a host recombinant cell by expression of the nucleotide sequence may be secreted or may be contained intracellularly depending on the sequence and/or the vector used. The coding sequences can be designed with signal sequences which direct secretion of the substance coding sequences through a particular prokaryotic or eukaryotic cell membrane.

CONSTRUCTS

The term "construct" - which is synonymous with terms such as "conjugate", "cassette" and "hybrid" - includes a nucleotide sequence encoding a polypeptide having the specific properties as defined herein for use according to the present invention directly or indirectly attached to a promoter. An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, such as the Sh1-intron or the ADH intron, intermediate the promoter and the nucleotide sequence of the present invention. The same is true for the term "fused" in relation to the present invention which includes direct or indirect attachment. In some cases, the terms do not cover the natural combination of the nucleotide sequence coding for the protein ordinarily associated with the wild type gene promoter and when they are both in their natural environment.

The construct may even contain or express a marker which allows for the selection of the genetic construct.

For some applications, preferably the construct comprises at least a nucleotide sequence of the present invention or a nucleotide sequence encoding a polypeptide having the specific properties as defined herein operably linked to a promoter.

ORGANISM

The term "organism" in relation to the present invention includes any organism that could comprise a nucleotide sequence according to the present invention or a nucleotide sequence encoding for a polypeptide having the specific properties as defined herein and/or products obtained therefrom.

The term "transgenic organism" in relation to the present invention includes any organism that comprises a nucleotide sequence coding for a polypeptide having the specific properties as defined herein and/or the products obtained therefrom, and/or wherein a promoter can allow expression of the nucleotide sequence coding for a polypeptide having the specific properties as defined herein within the organism. Preferably the nucleotide sequence is incorporated in the genome of the organism.

The term "transgenic organism" does not cover native nucleotide coding sequences in their natural environment when they are under the control of their native promoter which is also in its natural environment.

Therefore, the transgenic organism of the present invention includes an organism comprising any one of, or combinations of, a nucleotide sequence coding for a polypeptide having the specific properties as defined herein, constructs as defined herein, vectors as defined herein, plasmids as defined herein, cells as defined herein, or the products thereof. For example the transgenic organism can also comprise a nucleotide sequence coding for a polypeptide having the specific properties as defined herein under the control of a promoter not associated with a sequence encoding a lipid acyltransferase in nature.

TRANSFORMATION OF HOST CELLS/ORGANISM

Teachings on the transformation of prokaryotic hosts are well documented in the art, for example see Sambrook *et al* (Molecular Cloning: A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory Press). If a prokaryotic host is used then the nucleotide sequence may need to be suitably modified before transformation - such as by removal of introns.

Various methods are known for the transformation of *Bacillus* species.

SECRETION

Often, it is desirable for the polypeptide to be secreted from the expression host into the culture medium from where the enzyme may be more easily recovered. According to the present invention, the secretion leader sequence may be selected on the basis of the desired expression host. Hybrid signal sequences may also be used with the context of the present invention.

Typical examples of secretion leader sequences not associated with a nucleotide sequence encoding a lipid acyltransferase in nature are those originating from the fungal amyloglucosidase (AG) gene (*gla*A - both 18 and 24 amino acid versions e.g. from *Aspergillus*), the a-factor gene (yeasts e.g. *Saccharomyces, Kluyveromyces* and *Hansenula*) or the α-amylase gene (*Bacillus*).

DETECTION

A variety of protocols for detecting and measuring the expression of the amino acid sequence are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic and amino acid assays.

A number of companies such as Pharmacia Biotech (Piscataway, NJ), Promega (Madison, WI), and US Biochemical Corp (Cleveland, OH) supply commercial kits and protocols for these procedures.

Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US-A-3,817,837; US-A-3,850,752; US-A-3,939,350; US-A-3,996,345; US-A-4,277,437; US-A-4,275,149 and US-A-4,366,241.

Also, recombinant immunoglobulins may be produced as shown in US-A-4,816,567.

FUSION PROTEINS

In the method of the present invention the lipid acyltransferase may be produced as a fusion protein, for example to aid in extraction and purification thereof. Examples of fusion protein partners include glutathione-S-transferase (GST), 6xHis, GAL4 (DNA binding and/or transcriptional activation domains) and β-galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences. Preferably the fusion protein will not hinder the activity of the protein sequence.

Gene fusion expression systems in *E. coli* have been reviewed in Curr. Opin. Biotechnol. (1995) 6(5):501-6.

In another embodiment of the invention, the amino acid sequence of a polypeptide having the specific properties as defined herein may be ligated to a non-native sequence to encode a fusion protein. For example, for screening of peptide libraries for agents capable of affecting the substance activity, it may be useful to encode a chimeric substance expressing a non-native epitope that is recognised by a commercially available antibody.

The invention will now be described, by way of example only, with reference to the following Figures and Examples.

Figure 1 shows the amino acid sequence of a mutant *Aeromonas salmonicida* mature lipid acyltransferase (GCAT) with a mutation of Asn80Asp (notably, amino acid 80 is in the mature sequence) (SEQ ID 16);

Figure 2 shows an amino acid sequence (SEQ ID No. 1) a lipid acyl transferase from *Aeromonas hydrophila* (ATCC #7965);

Figure 3 shows a pfam00657 consensus sequence from database version 6 (SEQ ID No. 2);

Figure 4 shows an amino acid sequence (SEQ ID No. 3) obtained from the organism *Aeromonas hydrophila* (P10480; GI:121051);

Figure 5 shows an amino acid sequence (SEQ ID No. 4) obtained from the organism *Aeromonas salmonicida* (AAG098404; GI:9964017);

Figure 6 shows an amino acid sequence (SEQ ID No. 5) obtained from the organism *Streptomyces coelicolor* A3(2) (Genbank accession number NP_631558);

Figure 7 shows an amino acid sequence (SEQ ID No. 6) obtained from the organism *Streptomyces coelicolor* A3(2) (Genbank accession number: CAC42140);

Figure 8 shows an amino acid sequence (SEQ ID No. 7) obtained from the organism *Saccharomyces cerevisiae* (Genbank accession number P41734);

Figure 9 shows an amino acid sequence (SEQ ID No. 8) obtained from the organism *Ralstonia* (Genbank accession number: AL646052);

Figure 10 shows SEQ ID No. 9. Scoe1 NCBI protein accession code CAB39707.1 GI:4539178 conserved hypothetical protein [Streptomyces coelicolor A3(2)];

Figure 11 shows an amino acid shown as SEQ ID No. 10. Scoe2 NCBI protein accession code CAC01477.1 GI:9716139 conserved hypothetical protein [Streptomyces coelicolor A3(2)];

Figure 12 shows an amino acid sequence (SEQ ID No. 11) Scoe3 NCBI protein accession code CAB88833.1 GI:7635996 putative secreted protein. [Streptomyces coelicolor A3(2)];

Figure 13 shows an amino acid sequence (SEQ ID No. 12) Scoe4 NCBI protein accession code CAB89450.1 GI:7672261 putative secreted protein. [Streptomyces coelicolor A3(2)];

Figure 14 shows an amino acid sequence (SEQ ID No. 13) Scoe5 NCBI protein accession code CAB62724.1 GI:6562793 putative lipoprotein [Streptomyces coelicolor A3(2)];

Figure 15 shows an amino acid sequence (SEQ ID No. 14) Srim1 NCBI protein accession code AAK84028.1 GI:15082088 GDSL-lipase [Streptomyces rimosus];

Figure 16 shows an amino acid sequence (SEQ ID No. 15) of a lipid acyltransferase from *Aeromonas salmonicida* subsp. *Salmonicida* (ATCC#14174);

Figure 17 shows SEQ ID No. 19. Scoe1 NCBI protein accession code CAB39707.1 GI:4539178 conserved hypothetical protein [Streptomyces coelicolor A3(2)];

Figure 18 shows an amino acid sequence (SEQ ID No. 25) of the fusion construct used for mutagenesis of the *Aeromonas hydrophila* lipid acyltransferase gene. The underlined amino acids is a xylanase signal peptide;

Figure 19 shows a polypeptide sequence of a lipid acyltransferase enzyme from *Streptomyces* (SEQ ID No. 26);

Figure 20 shows a polypeptide sequence of a lipid acyltransferase enzyme from *Thermobifida* (SEQ ID No. 27);

Figure 21 shows a polypeptide sequence of a lipid acyltransferase enzyme from *Thermobifida* (SEQ ID No. 28);

Figure 22 shows a polypeptide of a lipid acyltransferase enzyme from *Corynebacterium efficiens* GDSx 300 amino acid_(SEQ ID No. 29);

Figure 23 shows a polypeptide of a lipid acyltransferase enzyme from *Novosphingobium aromaticivorans* GDSx 284 amino acid_(SEQ ID No. 30);

Figure 24 shows a polypeptide of a lipid acyltransferase enzyme from *Streptomyces coelicolor* GDSx 269 aa (SEQ ID No. 31);

Figure 25 shows a polypeptide of a lipid acyltransferase enzyme from *Streptomyces avermitilis* \ GDSx 269 amino acid (SEQ ID No. 32);

Figure 26 shows a polypeptide of a lipid acyltransferase enzyme from *Streptomyces* (SEQ ID No. 33);

Figure 27 shows an amino acid sequence (SEQ ID No. 34) obtained from the organism *Aeromonas hydrophila* (P10480; GI:121051) (notably, this is the mature sequence);

Figure 28 shows the amino acid sequence (SEQ ID No. 35) of a mutant *Aeromonas salmonicida* mature lipid acyltransferase (GCAT) (notably, this is the mature sequence);

Figure 29 shows a nucleotide sequence (SEQ ID No. 36) from *Streptomyces thermosacchari;*

Figure 30 shows an amino acid sequence (SEQ ID No. 37) from *Streptomyces thermosacchari;*

Figure 31 shows an amino acid sequence (SEQ ID No. 38) from *Thermobifida fusca*/GDSx 548 amino acid;

Figure 32 shows a nucleotide sequence (SEQ ID No. 39) from *Thermobifida fusca;*

Figure 33 shows an amino acid sequence (SEQ ID No. 40) from *Thermobifida fusca*/GDSx;

Figure 34 shows an amino acid sequence (SEQ ID No. 41) from *Corynebacterium efficiens*/GDSx 300 amino acid;

Figure 35 shows a nucleotide sequence (SEQ ID No. 42) from *Corynebacterium efficiens;*

Figure 36 shows an amino acid sequence (SEQ ID No. 43) from *S. coelicolorl* GDSx 268 amino acid;

Figure 37 shows a nucleotide sequence (SEQ ID No. 44) from *S. coelicolor;*

Figure 38 shows an amino acid sequence (SEQ ID No. 45) from *S. avermitilis;*

Figure 39 shows a nucleotide sequence (SEQ ID No. 46) from *S. avermitilis;*

Figure 40 shows an amino acid sequence (SEQ ID No. 47) from *Thermobifida fusca/GDSx;*

Figure 41 shows a nucleotide sequence (SEQ ID No. 48) from *Thermobifida fusca/GDSx;*

Figure 42 shows an alignment of the L131 and homologues from *S. avermitilis* and *T. fusca* illustrates that the conservation of the GDSx motif (GDSY in L131 and *S. avermitilis* and *T. fusca*), the GANDY box, which is either GGNDA or GGNDL, and the HPT block (considered to be the conserved catalytic histidine). These three conserved blocks are highlighted;

Figure 43 shows SEQ ID No 17 which is the amino acid sequence of a lipid acyltransferase from *Candida parapsilosis*;

Figure 44 shows SEQ ID No 18 which is the amino acid sequence of a lipid acyltransferase from *Candida parapsilosis*;

Figure 45 shows a ribbon representation of the 1IVN.PDB crystal structure which has glycerol in the active site. The Figure was made using the Deep View Swiss-PDB viewer;

Figure 46 shows 1IVN.PDB Crystal Structure – Side View using Deep View Swiss-PDB viewer, with glycerol in active site - residues within 10Å of active site glycerol are coloured black;

Figure 47 shows 1IVN.PDB Crystal Structure – Top View using Deep View Swiss-PDB viewer, with glycerol in active site – residues within 10Å of active site glycerol are coloured black;

Figure 48 shows alignment 1;

Figure 49 shows alignment 2;

Figures 50 and 51 show an alignment of 1IVN to P10480 (P10480 is the database sequence for *A. hydrophila* enzyme), this alignment was obtained from the PFAM database and used in the model building process; and

Figure 52 shows an alignment where P10480 is the database sequence for *Aeromonas hydrophila*. This sequence is used for the model construction and the site selection. Note that the full protein (SEQ ID No. 25) is depicted, the mature protein (equivalent to SEQ ID No. 34) starts at residue 19. A. sal is *Aeromonas salmonicida* (SEQ ID No. 4) GDSX lipase, A. hyd is *Aeromonas hydrophila* (SEQ ID No. 34) GDSX lipase. The consensus sequence contains a * at the position of a difference between the listed sequences.

Figure 53 shows a gene construct used in Example 1;

Figure 54 shows a codon optimised gene construct (no. 052907) used in Example 1; and

Figure 55 shows the sequence of the XhoI insert containing the LAT-KLM3' precursor gene, the -35 and -10 boxes are underlined;

Figure 56 shows BML780-KLM3'CAP50 (comprising SEQ ID No. 16 – upper colony) and BML780 (the empty host strain – lower colony) after 48h growth at 37°C on 1% tributyrin agar;

Figure 57 shows a nucleotide sequence from *Aeromonas salmonicida* (SEQ ID No. 49) including the signal sequence (preLAT - positions 1 to 87);

Figure 58 shows a nucleotide sequence (SEQ ID No. 50) encoding a lipid acyl transferase according to the present invention obtained from the organism *Aeromonas hydrophila*;

Figure 59 shows a nucleotide sequence (SEQ ID No. 51) encoding a lipid acyl transferase according to the present invention obtained from the organism *Aeromonas salmonicida*;

Figure 60 shows a nucleotide sequence (SEQ ID No. 52) encoding a lipid acyl transferase according to the present invention obtained from the organism *Streptomyces coelicolor* A3(2) (Genbank accession number NC_003888.1:8327480..8328367);

Figure 61 shows a nucleotide sequence (SEQ ID No. 53) encoding a lipid acyl transferase according to the present invention obtained from the organism *Streptomyces coelicolor* A3(2) (Genbank accession number AL939131.1:265480..266367);

Figure 62 shows a nucleotide sequence (SEQ ID No. 54) encoding a lipid acyl transferase according to the present invention obtained from the organism *Saccharomyces cerevisiae* (Genbank accession number Z75034);

Figure 63 shows a nucleotide sequence (SEQ ID No. 55) encoding a lipid acyl transferase according to the present invention obtained from the organism *Ralstonia*;

Figure 64 shows a nucleotide sequence shown as SEQ ID No. 56 encoding NCBI protein accession code CAB39707.1 GI:4539178 conserved hypothetical protein [*Streptomyces coelicolor* A3(2)];

Figure 65 shows a nucleotide sequence shown as SEQ ID No. 57 encoding Scoe2 NCBI protein accession code CAC01477.1 GI:9716139 conserved hypothetical protein [*Streptomyces coelicolor* A3(2)];

Figure 66 shows a nucleotide sequence shown as SEQ ID No. 58 encoding Scoe3 NCBI protein accession code CAB88833.1 GI:7635996 putative secreted protein. [*Streptomyces coelicolor* A3(2)];

Figure 67 shows a nucleotide sequence shown as SEQ ID No. 59 encoding Scoe4 NCBI protein accession code CAB89450.1 GI:7672261 putative secreted protein. [*Streptomyces coelicolor* A3(2)];

Figure 68 shows a nucleotide sequence shown as SEQ ID No. 60, encoding Scoe5 NCBI protein accession code CAB62724.1 GI:6562793 putative lipoprotein [*Streptomyces coelicolor* A3(2)];

Figure 69 shows a nucleotide sequence shown as SEQ ID No. 61 encoding Srim1 NCBI protein accession code AAK84028.1 GI:15082088 GDSL-lipase [*Streptomyces rimosus*];

Figure 70 shows a nucleotide sequence (SEQ ID No. 62) encoding a lipid acyltransferase from *Aeromonas hydrophila* (ATCC #7965);

Figure 71 shows a nucleotide sequence (SEQ ID No 63) encoding a lipid acyltransferase from *Aeromonas salmonicida* subsp. *Salmonicida* (ATCC#14174); and

Figure 72 shows a nucleotide sequence (SEQ ID No. 24) encoding an enzyme from *Aeromonas hydrophila* including a xylanase signal peptide.

EXAMPLE 1

Expression of KLM3' in *Bacillus licheniformis*

A nucleotide sequence (SEQ ID No. 49) encoding a lipid acyltransferase (SEQ. ID No. 16, hereinafter KLM3') was expressed in *Bacillus licheniformis* as a fusion protein with the signal peptide of *B. licheniformis* [alpha]-amylase (LAT) (see FIGS. 53 and 54). For optimal expression in Bacillus, a codon optimized gene construct (no. 052907) was ordered at Geneart (Geneart AG, Regensburg, Germany).

Construct no. 052907 contains an incomplete LAT promoter (only the –10 sequence) in front of the LAT-KLM3' precursor gene and the LAT transcription (Tlat) downstream of the LAT-KLM3' precursor gene (see FIGS 53 and 55). To create a *Xho*I fragment that contains the LAT-KLM3' precursor gene flanked by the complete LAT promoter at the 5' end and the LAT terminator at the 3' end, a PCR (polymerase chain reaction) amplification was performed with the primers Plat5XhoI_FW and EBS2XhoI_RV and gene construct 052907 as template.

Plat5XhoI_FW:

ccccgctcgaggcttttcttttggaagaaaatatagggaaaatggtacttgttaaaaattc
ggaatatttatacaatatcatatgtttcacattgaaagggg

EBS2XhoI_RV: tggaatctcgaggttttatcctttaccttgtctcc

PCR was performed on a thermocycler with Phusion High Fidelity DNA polymerase (Finnzymes OY, Espoo, Finland) according to the instructions of the manufacturer (annealing temperature of 55[deg.] C.).

The resulting PCR fragment was digested with restriction enzyme *Xho*I and ligated with T4 DNA ligase into *Xho*I digested pICatH according to the instructions of the supplier (Invitrogen, Carlsbad, Calif. USA).

The ligation mixture was transformed into *B. subtilis* strain SC6.1 as described in U.S. Patent Application US20020182734 (International Publication WO 02/14490). The

sequence of the *Xho*I insert containing the LAT-KLM3' precursor gene was confirmed by DNA sequencing (BaseClear, Leiden, The Netherlands) and one of the correct plasmid clones was designated pICatH-KLM3'(ori1) (Figure 53). pICatH-KLM3'(ori1) was transformed into *B. licheniformis* strain BML780 (a derivative of BRA7 and BML612, see WO2005111203) at the permissive temperature (37[deg.] C.).

One neomycin resistant (neoR) and chloramphenicol resistant (CmR) transformant was selected and designated BML780(pICatH-KLM3'(ori1)). The plasmid in BML780(pICatH-KLM3'(ori1)) was integrated into the catH region on the *B. licheniformis* genome by growing the strain at a non-permissive temperature (50[deg.] C) in medium with 5 [mu]g/ml chloramphenicol. One CmR resistant clone was selected and designated BML780-pICatH-KLM3'(ori1). BML780-pICatH- KLM3'(ori1) was grown again at the permissive temperature for several generations without antibiotics to loop-out vector sequences and then one neomycin sensitive (neoS), CmR clone was selected. In this clone, vector sequences of pICatH on the chromosome are excised (including the neomycin resistance gene) and only the catH - LATKLM3' cassette is left. Next, the catH - LATKLM3' cassette on the chromosome was amplified by growing the strain in/on media with increasing concentrations of chloramphenicol. After various rounds of amplification, one clone (resistant against 50 [mu]g/ml chloramphenicol) was selected and designated BML780-KLM3'CAP50. To verify KLM3'expression, BML780-KLM3'CAP50 and BML780 (the empty host strain) were grown for 48h at 37 [deg.] C on a Heart Infusion (Bacto) agar plate with 1% tributyrin. A clearing zone, indicative for lipid acyltransferase activity, was clearly visible around the colony of BML780-KLM3'CAP50 but not around the host strain BML780 (see Figure 56). This result shows that a substantial amount of KLM3' is expressed in *B. licheniformis* strain BML780-KLM3'CAP50 and that these KLM3' molecules are functional.

COMPARATIVE EXAMPLE 1

Vector construct

The plasmid construct is pCS32new N80D, which is a pCCmini derivative carrying the sequence encoding the mature form of the native *Aeromonas salmonicida* Glycerophospholipid-cholesterol acyltransferase with a Asn to Asp substitution at

position 80 (KLM3'), under control of the p32 promoter and with a CGTase signal sequence.

The host strain used for the expression, is in the *bacillus subtilis* OS21ΔAprE strain

The expression level is measured as transferase activity, expressed as % cholesterol esterified, calculated from the difference in free cholesterol in the reference sample and free cholesterol in the enzyme sample in reactions with PC (T_{PC}) as donor and cholesterol as acceptor molecule.

Culture conditions

5 ml of LB broth (Casein enzymatic digest, 10 g/l; low-sodium Yeast extract, 5 g/l; Sodium Chloride, 5 g/l; Inert tableting aids, 2 g/l) supplemented with 50 mg/l kanamycin, was inoculated with a single colony and incubated at 30 °C for 6 hours at 205 rpm. 0.7 ml of this culture was used to inoculate 50 ml of SAS media (K_2HPO_4, 10 g/l; MOPS (3-morpholinopropane sulfonic acid), 40 g/l; Sodium Chloride, 5 g/l; Antifoam (Sin 260), 5 drops/l; Soy flour degreased, 20 g/l; Biospringer 106 (100 % dw YE), 20 g/l) supplemented with 50 mg/l kanamycin and a solution of high maltose starch hydrolysates (60 g/l). Incubation was continued for 40 hours at 30 °C and 180 rpm before the culture supernatant was separated by centrifugation at 19000 rpm for 30 min. The supernatant was transferred into a clean tube and directly used for transferase activity measurement.

Preparation of substrates and enzymatic reaction

PC (Avanti Polar Lipids #441601) and cholesterol (Sigma C8503) was scaled in the ratio 9:1, dissolved in chloroform, and evaporated to dryness.

The substrate was prepared by dispersion of 3% PC:Cholesterol 9:1 in 50 mM Hepes buffer pH 7.

0.250 ml substrate solution was transferred into a 3 ml glass tube with screw lid. 0.025 ml culture supernatant was added and the mixture was incubated at 40 °C for 2 hours. A reference sample with water instead of enzyme was also prepared. Heating the reaction mixture in a boiling water bath for 10 minutes stopped the enzyme reaction. 2 ml of 99% ethanol was added to the reaction mixture before submitted to cholesterol assay analysis.

Cholesterol assay

100 µl substrate containing 1.4 U/ml Cholesterol oxidase(SERVA Electrophoresis GmbH cat. No 17109), 0.4 mg/ml ABTS (Sigma A-1888), 6 U/ml Peroxidase (Sigma 6782) in 0.1 M Tris-HCl, pH 6.6 and 0.5 % Triton™ X-100 (Sigma X-100) was incubated at 37°C for 5 minutes before 5 µl enzyme reaction sample was added and mixed. The reaction mixture was incubated for further 5 minutes and OD_{405} was measured. The content of cholesterol was calculated from the analyses of standard solutions of cholesterol containing 0.4 mg/ml, 0.3 mg/ml, 0.20 mg/ml, 0.1 mg/ml, 0.05 mg/ml, and 0 mg/ml cholesterol in 99 % EtOH.

Results

The table shows the average of 8 separate expression cultures

Strain	T_{PC}[a]
OS21ΔAprE[pCS52new]	74.2 ±10.1[b]

[a] T_{PC} is the transferase activity, expressed as % cholesterol esterified, calculated from the difference in free cholesterol in the reference sample and free cholesterol in the enzyme sample in reactions with PC as donor molecule and cholesterol as acceptor molecule.

[b] Average of 8 separate expression cultures

Various modifications and variations of the described methods and system of the present invention will be apparent to those skilled in the art. Although the present invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in biochemistry and biotechnology or related fields are intended to be within the scope of the following claims.

CLAIMS:

1. A method for the production of a lipid acyltransferase comprising the steps of:

(i) providing a *Bacillus lichenifomis* cell,

(ii) transforming the *Bacillus lichenifomis* cell with a heterologous nucleotide sequence encoding a lipid acyltransferase shown as SEQ ID NO. 49, and

(iii) expressing the lipid acyltransferase in the cell under the control of a promoter sequence.

2. The method according to claim 1, wherein the promoter sequence is not natively associated with the nucleotide sequence encoding a lipid acyltransferase.

3. The method according to claim 1 or claim 2, wherein a nucleotide sequence encoding a signal peptide is operably linked to said heterologous nucleotide sequence encoding a lipid acyltransferase.

4. The method according to any one of claims 1 to 3, wherein said method comprises the additional step of isolating the lipid acyltransferase.

5. The method according to any one of claims 1 to 4, wherein said promoter sequence is homologous to the host cell.

6. The method according to any one of claims 1 to 5, wherein said promoter sequence is selected from the group consisting of an α-amylase promoter sequence, a protease promoter sequence, a subtilisin promoter sequence, a glutamic acid-specific protease promoter sequence and a levansucrase promoter sequence.

7. The method according to any one of claims 1 to 5, wherein said promoter sequence is an α-amylase promoter sequence.

8. A *Bacillus licheniformis* host cell comprising a nucleotide sequence encoding a heterologous lipid acyltransferase shown as SEQ ID NO. 49.

9. Use of a *Bacillus licheniformis* host cell in the production of a heterologous lipid acyltransferase encoded by a nucleotide sequence shown as SEQ ID NO. 49.

10. The use according to claim 9, wherein there is an increased expression of the lipid acyltransferase as compared with expression of the lipid acyltransferase in a *B. subtilis* host cell.

11. An expression vector comprising a nucleotide sequence encoding a lipid acyltransferase shown as SEQ ID NO. 49 operably linked to a promoter sequence homologous to *B licheniformis.*

12. The expression vector according to claim 11, wherein the promoter is not natively associated with the nucleotide sequence encoding a lipid acyltransferase.

13. The expression vector according to claim 11 or claim 12, wherein a nucleotide sequence encoding a signal peptide is operably linked to a heterologous nucleotide sequence encoding the lipid acyltransferase.

14. The expression vector according to any one of claims 11 to 13, wherein said promoter sequence is selected from the group consisting of an α-amylase promoter sequence, a protease promoter sequence, a subtilisin promoter sequence, a glutamic acid-specific protease promoter sequence and a levansucrase promoter sequence.

15. An expression vector according to any one of claims 11 to 13, wherein said promoter sequence is an α-amylase promoter sequence.

16. A lipid acyltransferase obtained by the method of any one of claims 1 to 7.

FIGURE 1 (SEQ ID No. 16)

```
  1  ADTRPAFSRI VMFGDSLSDT GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTKQFPGLT
 61  IANEAEGGAT AVAYNKISWD PKYQVINNLD YEVTQFLQKD SFKPDDLVIL WVGANDYLAY
121  GWNTEQDAKR VRDAISDAAN RMVLNGAKQI LLFNLPDLGQ NPSARSQKVV EAVSHVSAYH
181  NKLLLNLARQ LAPTGMVKLF EIDKQFAEML RDPQNFGLSD VENPCYDGGY VWKPFATRSV
241  STDRQLSAFS PQERLAIAGN PLLAQAVASP MARRSASPLN CEGKMFWDQV HPTTVVHAAL
301  SERAATFIET QYEFLAHG
```

FIGURE 2 (SEQ ID No. 1)

```
  1  MKKWFVCLLG LVALTVQAAD SRPAFSRIVM FGDSLSDTGK MYSKMRGYLP
 51  SSPPYYEGRF SNGPVWLEQL TKQFPGLTIA NEAEGGATAV AYNKISWNPK
101  YQVINNLDYE VTQFLQKDSF KPDDLVILWV GANDYLAYGW NTEQDAKRVR
151  DAISDAANRM VLNGAKQILL FNLPDLGQNP SARSQKVVEA VSHVSAYHNQ
201  LLLNLARQLA PTGMVKLFEI DKQFAEMLRD PQNFGLSDVE NPCYDGGYVW
251  KPFATRSVST DRQLSAFSPQ ERLAIAGNPL LAQAVASPMA RRSASPLNCE
301  GKMFWDQVHP TTVVHAALSE RAATFIANQY EFLAH*
```

FIGURE 3 (SEQ ID No. 2)

```
  1 ivafGDSlTd geayygdsdg ggwgagladr Ltallrlrar prgvdvfnrg isGrtsdGrl
 61 ivDalvallF laqslglpnL pPYLsgdflr GANFAsagAt Ilptsgpfli QvqFkdfksq
121 vlelrqalgl lqellrllpv ldakspdlvt imiGtNDlit saffgpkste sdrnvsvpef
181 kdnlrqlikr Lrsnngarii vlitlvilnl gplGClPlkl alalassknv dasgclerln
241 eavadfneal relaiskled qlrkdglpdv kgadvpyvDl ysifqdldgi qnpsayvyGF
301 ettkaCCGyG gryNynrvCG naglcnvtak aCnpssylls flfwDgfHps ekGykavAea
361 l
```

FIGURE 4 (SEQ ID No. 3)

```
  1 mkkwfvcllg lvaltvqaad srpafsrivm fgdslsdtgk myskmrgylp ssppyyegrf
 61 sngpvwleql tnefpgltia neaeggptav aynkiswnpk yqvinnldye vtqflqkdsf
121 kpddlvilwv gandylaygw nteqdakrvr daisdaanrm vlngakeill fnlpdlgqnp
181 sarsqkvvea ashvsayhnq lllnlarqla ptgmvklfei dkqfaemlrd pqnfglsdqr
241 nacyggsyvw kpfasrsast dsqlsafnpq erlaiagnpl laqavaspma arsastlnce
301 gkmfwdqvhp ttvvhaalse paatfiesqy eflah
```

FIGURE 5 (SEQ ID No. 4)

```
  1 mkkwfvcllg lialtvqaad trpafsrivm fgdslsdtgk myskmrgylp ssppyyegrf
 61 sngpvwleql tkqfpgltia neaeggatav aynkiswnpk yqvynnldye vtqflqkdsf
121 kpddlvilwv gandylaygw nteqdakrvr daisdaanrm vlngakqill fnlpdlgqnp
181 sarsqkvvea vshvsayhnk lllnlarqla ptgmvklfei dkqfaemlrd pqnfglsdve
241 npcydggyvw kpfatrsvst drqlsafspq erlaiagnpl laqavaspma rrsasplnce
301 gkmfwdqvhp ttvvhaalse raatfietqy eflahg
```

FIGURE 6 (SEQ ID No. 5)

```
  1 mpkpalrrvm tatvaavgtl algltdatah aapaqatptl dyvalgdsys agsgvlpvdp
 61 anllclrsta nyphviadtt garltdvtcg aaqtadftra qypgvapqld algtgtdlvt
121 ltiggndnst finaitacgt agvlsggkgs pckdrhgtsf ddeieantyp alkeallgvr
181 arapharvaa lgypwitpat adpscflklp laagdvpylr aiqahlndav rraaeetgat
241 yvdfsgvsdg hdaceapgtr wiepllfghs lvpvhpnalg errmaehtmd vlgld
```

FIGURE 7 (SEQ ID No. 6)

```
  1 mpkpalrrvm tatvaavgtl algltdatah aapaqatptl dyvalgdsys agsgvlpvdp
 61 anllclrsta nyphviadtt garltdvtcg aaqtadftra qypgvapqld algtgtdlvt
121 ltiggndnst finaitacgt agvlsggkgs pckdrhgtsf ddeieantyp alkeallgvr
181 arapharvaa lgypwitpat adpscflklp laagdvpylr aiqahlndav rraaeetgat
241 yvdfsgvsdg hdaceapgtr wiepllfghs lvpvhpnalg errmaehtmd vlgld
```

FIGURE 8 (SEQ ID No. 7)

```
  1 mdyekfllfg dsitefafnt rpiedgkdqy algaalvney trkmdilqrg fkgytsrwal
 61 kilpeilkhe snivmatifl gandacsagp qsvplpefid nirqmvslmk syhirpiiig
121 pglvdrekwe kekseeialg yfrtnenfai ysdalaklan eekvpfvaln kafqqeggda
181 wqqlltdglh fsgkgykifh dellkvietf ypqyhpknmq yklkdwrdvl ddgsnims
```

FIGURE 9 (SEQ ID No. 8)

```
          10         20         30         40         50         60
          |          |          |          |          |          |
MNLRQWMGAA TAALALGLAA CGGGGTDQSG NPNVAKVQRM VVFGDSLSDI GTYTPVAQAV

          70         80         90         100        110        120
          |          |          |          |          |          |
GGGKFTTNPG PIWAETVAAQ LGVTLTPAVM GYATSVQNCP KAGCFDYAQG GSRVTDPNGI

          130        140        150        160        170        180
          |          |          |          |          |          |
GHNGGAGALT YPVQQQLANF YAASNNTFNG NNDVVFVLAG SNDIFFWTTA AATSGSGVTP

          190        200        210        220        230        240
          |          |          |          |          |          |
AIATAQVQQA ATDLVGYVKD MIAKGATQVY VFNLPDSSLT PDGVASGTTG QALLHALVGT

          250        260        270        280        290        300
          |          |          |          |          |          |
FNTTLQSGLA GTSARIIDFN AQLTAAIQNG ASFGFANTSA RACDATKINA LVPSAGGSSL

          310        320        330        340
          |          |          |          |
FCSANTLVAS GADQSYLFAD GVHPTTAGHR LIASNVLARL LADNVAH
```

FIGURE 10 (SEQ ID No. 9)

```
  1 migsyvavgd sftegvgdpg pdgafvgwad rlavlladrr pegdftytnl avrgrlldqi
 61 vaeqvprvvg lapdlvsfaa ggndiirpgt dpdevaerfe lavaaltaaa gtvlvttgfd
121 trgvpvlkhl rgkiatyngh vraiadrygc pvldlwslrs vqdrrawdad rlhlspeght
181 rvalragqal glrvpadpdq pwpplpprgt ldvrrddvhw areylvpwig rrlrgessgd
241 hvtakgtlsp daiktriaav a
```

FIGURE 11 (SEQ ID No. 10)

```
  1 mqtnpaytsl vavgdsfteg msdllpdgsy rgwadllatr maarspgfry anlavrgkli
 61 gqivdeqvdv aaamgadvit lvgglndtlr pkcdmarvrd lltqaverla phceqlvlmr
121 spgrqgpvle rfrprmealf aviddlagrh gavvvdlyga qsladprmwd vdrlhltaeg
181 hrrvaeavwq slghepedpe whapipatpp pgwvtrrtad vrfarqhllp wigrrltgrs
241 sgdglpakrp dllpyedpar
```

FIGURE 12 (SEQ ID No. 11)

```
  1 mtrgrdggag apptkhrall aaivtlivai saaiyagasa ddgsrdhalq aggrlprgda
 61 apastgawvg awatapaaae pgtettglag rsvrnvvhts vggtgaritl snlygqsplt
121 vthasialaa gpdtaaaiad tmrrltfggs arviipaggq vmsdtarlai pyganvlvtt
181 yspipsgpvt yhpqarqtsy ladgdrtadv tavayttptp ywryltaldv lsheadgtvv
241 afgdsitdga rsqsdanhrw tdvlaarlhe aagdgrdtpr ysvvnegisg nrlltsrpgr
301 padnpsglsr fqrdvlertn vkavvvvlgv ndvlnspela drdailtglr tlvdraharg
361 lrvvgatitp fggyggytea retmrqevne eirsgrvfdt vvdfdkalrd pydprrmrsd
421 ydsgdhlhpg dkgyarmgav idlaalkgaa pvka
```

FIGURE 13 (SEQ ID No. 12)

```
  1 mtsmsrarva rriaagaayg gggiglagaa avglvvaevq larrrvgvgt ptrvpnaqgl
 61 yggtlptagd pplrlmmlgd staagqgvhr agqtpgalla sglaavaerp vrlgsvaqpg
121 acsddldrqv alvlaepdrv pdicvimvga ndvthrmpat rsvrhlssav rrlrtagaev
181 vvgtcpdlgt iervrqplrw larrasrqla aaqtigaveq ggrtvslgdl lgpefaqnpr
241 elfgpdnyhp saegyataam avlpsvcaal glwpadeehp dalrregflp varaaaeaas
301 eagtevaaam ptgprgpwal lkrrrrrrvs eaepsspsgv
```

FIGURE 14 (SEQ ID No. 13)

```
  1 mgrgtdqrtr ygrrrarval aaltaavlgv gvagcdsvgg dspapsgsps krtrtapawd
 61 tspasvaavg dsitrgfdac avlsdcpevs watgssakvd slavrllgka daaehswnya
121 vtgarmadlt aqvtraaqre pelvavmaga ndacrsttsa mtpvadfraq feeamatlrk
181 klpkaqvyvs sipdlkrlws qgrtnplgkq vwklglcpsm lgdadsldsa atlrrntvrd
241 rvadynevlr evcakdrrcr sddgavhefr fgtdqlshwd wfhpsvdgqa rlaeiayrav
301 taknp
```

FIGURE 15 (SEQ ID No. 14)

```
  1 mrlsrraata sallltpala lfgasaavsa priqatdyva lgdsyssgvg agsydsssgs
 61 ckrstksypa lwaashtgtr fnftacsgar tgdvlakqlt pvnsgtdlvs itiggndagf
121 adtmttcnlq gesaclaria karayiqqtl paqldqvyda idsrapaaqv vvlgyprfyk
181 lggscavgls eksraainaa addinavtak raadhgfafg dvnttfaghe lcsgapwlhs
241 vtlpvensyh ptangqskgy lpvlnsat
```

FIGURE 16 (SEQ ID No. 15)

```
  1 MKKWFVCLLG LIALTVQAAD TRPAFSRIVM FGDSLSDTGK MYSKMRGYLP
 51 SSPPYYEGRF SNGPVWLEQL TKQFPGLTIA NEAEGGATAV AYNKISWNPK
101 YQVINNLDYE VTQFLQKDSF KPDDLVILWV GANDYLAYGW NTEQDAKRVR
151 DAISDAANRM VLNGAKQILL FNLPDLGQNP SARSQKVVEA VSHVSAYHNK
201 LLLNLARQLA PTGMVKLFEI DKQFAEMLRD PQNFGLSDVE NPCYDGGYVW
251 KPFATRSVST DRQLSAFSPQ ERLAIAGNPL LAQAVASPMA RRSASPLNCE
301 GKMFWDQVHP TTVVHAALSE RAATFIETQY EFLAHG*
```

FIGURE 17 (SEQ ID No. 19)

```
  1 migsyvavgd sftegvgdpg pdgafvgwad rlavlladrr pegdftytnl avrgrlldqi
 61 vaeqvprvvg lapdlvsfaa ggndiirpgt dpdevaerfe lavaaltaaa gtvlvttgfd
121 trgvpvlkhl rgkiatyngh vraiadrygc pvldlwslrs vqdrrawdad rlhlspeght
181 rvalragqal glrvpadpdq pwpplpprgt ldvrrddvhw areylvpwig rrlrgessgd
241 hvtakgtlsp daiktriaav a
```

FIGURE 18 (SEQ ID No. 25)

```
  1 MFKFKKNFLV GLSAALMSIS LFSATASAAS ADSRPAFSRI VMFGDSLSDT
 51 GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTKQFPGLT IANEAEGGAT
101 AVAYNKISWN PKYQVINNLD YEVTQFLQKD SFKPDDLVIL WVGANDYLAY
151 GWNTEQDAKR VRDAISDAAN RMVLNGAKQI LLFNLPDLGQ NPSARSQKVV
201 EAVSHVSAYH NQLLLNLARQ LAPTGMVKLF EIDKQFAEML RDPQNFGLSD
251 VENPCYDGGY VWKPFATRSV STDRQLSAFS PQERLAIAGN PLLAQAVASP
301 MARRSASPLN CEGKMFWDQV HPTTVVHAAL SERAATFIAN QYEFLAH**
```

FIGURE 19 (SEQ ID NO. 26)

```
MRLTRSLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSGNGAGSYIDSSGDCHRSN
NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSITIGGNDAGFADAMTT
CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC
LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE
SYHPTSTGHQSGYLPVLNANSST
```

FIGURE 20 (SEQ ID No. 27)

ZP_00058717

```
   1 mlphpagerg evgaffallv gtpqdrrlrl echetrplrg rcgcgerrvp pltlpgdgvl
  61 cttsstrdae tvwrkhlqpr pdggfrphlg vgcllagqgs pgvlwcgreg crfevcrrdt
 121 pglsrtrngd ssppfragws lppkcgeisq sarktpavpr ysllrtdrpd gprgrfvgsg
 181 praatrrrlf lgipalvlvt altlvlavpt gretlwrmwc eatqdwclgv pvdsrgqpae
 241 dgefllllspv qaatwgnyya lgdsyssgdg ardyypgtav kggcwrsana ypelvaeayd
 301 faghlsflac sgqrgyamld aidevgsqld wnsphtslvt igiggndlgf stvlktcmvr
 361 vplldskact dqedairkrm akfettfeel isevrtrapd arilvvgypr ifpeeptgay
 421 ytltasnqrw lnetiqefnq qlaeavavhd eeiaasggvg svefvdvyha ldgheigsde
 481 pwvngvqlrd latgvtvdrs tfhpnaaghr avgervieqi etgpgrplya tfavvagatv
 541 dtlagevg
```

FIGURE 21 (SEQ ID No. 28)

```
   1 mgsgpraatr rrlflgipal vlvtaltlvl avptgretlw rmwceatqdw clgvpvdsrg
  61 qpaedgefll lspvqaatwg nyyalgdsys sgdgardyyp gtavkggcwr sanaypelva
 121 eaydfaghls flacsgqrgy amldaidevg sqldwnspht slvtigiggn dlgfstvlkt
 181 cmvrvpllds kactdqedai rkrmakfett feelisevrt rapdarilvv gyprifpeep
 241 tgayytltas nqrwlnetiq efnqqlaeav avhdeeiaas ggvgsvefvd vyhaldghei
 301 gsdepwvngv qlrdlatgvt vdrstfhpna aghravgerv ieqietgpgr plyatfavva
 361 gatvdtlage vg
```

FIGURE 22 (SEQ ID No. 29)

```
  1 mrttviaasa llllagcadg areetagapp gessggiree gaeastsitd vyialgdsya
 61 amggrdqplr gepfclrssg nypellhaev tdltcqgavt gdlleprtlg ertlpaqvda
121 ltedttlvtl siggndlgfg evagcireri agenaddcvd llgetigeql dqlppqldrv
181 heairdragd aqvvvtgylp lvsagdcpel gdvseadrrw aveltgqine tvreaaerhd
241 alfvlpddad ehtscappqq rwadiqgqqt dayplhptsa gheamaaavr dalglepvqp
```

FIGURE 23 (SEQ ID No. 30)

ZP_00094165

```
  1 mgqvklfarr capvllalag lapaatvare aplaegaryv algssfaagp gvgpnapgsp
 61 ercgrgtlny phllaealkl dlvdatcsga tthhvlgpwn evppqidsvn gdtrlvtlti
121 ggndvsfvgn ifaaacekma spdprcgkwr eiteeewqad eermrsivrq iharaplarv
181 vvvdyitvlp psgtcaamai spdrlaqsrs aakrlarita rvareegasl lkfshisrrh
241 hpcsakpwsn glsapaddgi pvhpnrlgha eaaaalvklv klmk //
```

FIGURE 24 (SEQ ID No. 31)

NP_625998.

```
      1 mrrfrlvgfl sslvlaagaa ltgaataqaa qpaaadgyva lgdsyssgvg agsyissgd
     61 ckrstkahpy lwaaahspst fdftacsgar tgdvlsgqlg plssgtglvs isiggndagf
    121 adtmttcvlq sessclsria taeayvdstl pgkldgvysa isdkapnahv vvigyprfyk
    181 lgttciglse tkrtainkas dhlntvlaqr aaahgftfgd vrttftghel csgspwlhsv
    241 nwlnigesyh ptaagqsggy lpvlngaa
//
```

FIGURE 25 (SEQ ID No. 32)

NP_827753.

```
      1 mrrsritayv tslllavgca ltgaataqas paaaatgyva lgdsyssgvg agsylsssgd
     61 ckrsskaypy lwqaahspss fsfmacsgar tgdvlanqlg tlnsstglvs ltiggndagf
    121 sdvmttcvlq sdsaclsrin takayvdstl pgqldsvyta istkapsahv avlgyprfyk
    181 lggsclagls etkrsainda adylnsaiak raadhgftfg dvkstftghe icssstwlhs
    241 ldllnigqsy hptaagqsgg ylpvmnsva
//
```

FIGURE 26 (SEQ ID No. 33)

```
MRLTRSLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSGNGAGSYIDSSGDCHRSN
NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSITIGGNDAGFADAMTT
CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC
LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE
SYHPTSTGHQSGYLPVLNANSST
```

FIGURE 27 (SEQ ID No. 34)

ADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLPSSPPYYEGRFSNGPVWLEQLTNEFPGLTIANEAEGGPT
AVAYNKISWNPKYQVINNLDYEVTQFLQKDSFKPDDLVILWVGANDYLAYGWNTEQDAKRVRDAISDAAN
RMVLNGAKEILLFNLPDLGQNPSARSQKVVEAASHVSAYHNQLLLNLARQLAPTGMVKLFEIDKQFAEML
RDPQNFGLSDQRNACYGGSYVWKPFASRSASTDSQLSAFNPQERLAIAGNPLLAQAVASPMAARSASTLN
CE
GKMFWDQVHPTTVVHAALSEPAATFIESQYEFLAH

FIGURE 28 (SEQ ID No. 35)

```
  1  ADTRPAFSRI VMFGDSLSDT GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTKQFPGLT
 61  IANEAEGGAT AVAYNKISWN PKYQVINNLD YEVTQFLQKD SFKPDDLVIL WVGANDYLAY
121  GWNTEQDAKR VRDAISDAAN RMVLNGAKQI LLFNLPDLGQ NPSARSQKVV EAVSHVSAYH
181  NKLLLNLARQ LAPTGMVKLF EIDKQFAEML RDPQNFGLSD VENPCYDGGY VWKPFATRSV
241  STDRQLSAFS PQERLAIAGN PLLAQAVASP MARRSASPLN CEGKMFWDQV HPTTVVHAAL
301  SERAATFIET QYEFLAHG
```

FIGURE 29 (SEQ ID No. 36)

ACAGGCCGATGCACGGAACCGTACCTTTCCGCAGTGAAGCGCTCTCCCCCCATCGTTCGC
CGGGACTTCATCCGCGATTTTGGCATGAACACTTCCTTCAACGCGCGTAGCTTGCTACAA
GTGCGGCAGCAGACCCGCTCGTTGGAGGCTCAGTGAGATTGACCCGATCCCTGTCGGCCG
CATCCGTCATCGTCTTCGCCCTGCTGCTCGCGCTGCTGGGCATCAGCCCGGCCCAGGCAG
CCGGCCCGGCCTATGTGGCCCTGGGGGATTCCTATTCCTCGGGCAACGGCGCCGGAAGTT
ACATCGATTCGAGCGGTGACTGTCACCGCAGCAACAACGCGTACCCCGCCCGCTGGGCGG
CGGCCAACGCACCGTCCTCCTTCACCTTCGCGGCCTGCTCGGGAGCGGTGACCACGGATG
TGATCAACAATCAGCTGGGCGCCCTCAACGCGTCCACCGGCCTGGTGAGCATCACCATCG
GCGGCAATGACGCGGGCTTCGCGGACGCGATGACCACCTGCGTCACCAGCTCGGACAGCA
CCTGCCTCAACCGGCTGGCCACCGCCACCAACTACATCAACACCACCCTGCTCGCCCGGC
TCGACGCGGTCTACAGCCAGATCAAGGCCCGTGCCCCCAACGCCCGCGTGGTCGTCCTCG
GCTACCCGCGCATGTACCTGGCCTCGAACCCCTGGTACTGCCTGGGCCTGAGCAACACCA
AGCGCGCGGCCATCAACACCACCGCCGACACCCTCAACTCGGTGATCTCCTCCCGGGCCA
CCGCCCACGGATTCCGATTCGGCGATGTCCGCCCGACCTTCAACAACCACGAACTGTTCT
TCGGCAACGACTGGCTGCACTCACTCACCCTGCCGGTGTGGGAGTCGTACCACCCCACCA
GCACGGGCCATCAGAGCGGCTATCTGCCGGTCCTCAACGCCAACAGCTCGACCTGATCAA
CGCACGGCCGTGCCCGCCCCGCGCGTCACGCTCGGCGCGGGCGCCGCAGCGCGTTGATCA
GCCCACAGTGCCGGTGACGGTCCCACCGTCACGGTCGAGGGTGTACGTCACGGTGGCGCC
GCTCCAGAAGTGGAACGTCAGCAGGACCGTGGAGCCGTCCCTGACCTCGTCGAAGAACTC
CGGGGTCAGCGTGATCACCCCTCCCCCGTAGCCGGGGGCGAAGGCGGCGCCGAACTCCTT
GTAGGACGTCCAGTCGTGCGGCCCGGCGTTGCCACCGTCCGCGTAGACCGCTTCCATGGT
CGCCAGCCGGTCCCCGCGGAACTCGGTGGGGATGTCCGTGCCCAAGGTGGTCCCGGTGGT
GTCCGAGAGCACCGGGGGCTCGTACCGGATGATGTGCAGATCCAAAGAATT

FIGURE 30 (SEQ ID NO. 37):

MRLTRSLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSGNGAGSYIDSSGDCHRSN
NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSITIGGNDAGFADAMTT
CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC
LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE
SYHPTSTGHQSGYLPVLNANSST

FIGURE 31 (SEQ ID No. 38)

```
  1 mlphpagerg evgaffallv gtpqdrrlrl echetrplrg rcgcgerrvp pltlpgdgvl
 61 cttsstrdae tvwrkhlqpr pdggfrphlg vgcllagqgs pgvlwcgreg crfevcrrdt
121 pglsrtrngd ssppfragws lppkcgeisq sarktpavpr ysllrtdrpd gprgrfvgsg
181 praatrrrlf lgipalvlvt altlvlavpt gretlwrmwc eatqdwclgv pvdsrgqpae
241 dgefllllspv qaatwgnyya lgdsyssgdg ardyypgtav kggcwrsana ypelvaeayd
301 faghlsflac sgqrgyamld aidevgsqld wnsphtslvt igiggndlgf stvlktcmvr
361 vplldskact dqedairkrm akfettfeel isevrtrapd arilvvgypr ifpeeptgay
421 ytltasnqrw lnetiqefnq qlaeavavhd eeiaasggvg svefvdvyha ldgheigsde
481 pwvngvqlrd latgvtvdrs tfhpnaaghr avgervieqi etgpgrplya tfavvagatv
541 dtlagevg
```

FIGURE 32 (SEQ ID No. 39)

```
   1 ggtggtgaac cagaacaccc ggtcgtcggc gtgggcgtcc aggtgcaggt gcaggttctt
  61 caactgctcc agcaggatgc cgccgtggcc gtgcacgatg gccttgggca ggcctgtggt
 121 ccccgacgag tacagcaccc atagcggatg gtcgaacggc agcggggtga actccagttc
 181 cgcgccttcg cccgcggctt cgaactccgc ccaggacagg gtgtcggcga cagggccgca
 241 gcccaggtac ggcaggacga cggtgtgctg caggctgggc atgccgtcgc gcagggcttt
 301 gagcacgtca cggcggtcga agtccttacc gccgtagcgg tagccgtcca cggccagcag
 361 cactttcggt tcgatctgcg cgaaccggtc gaggacgctg cgcaccccga agtcgggggа
 421 acaggacgac caggtcgcac cgatcgcggc gcaggcgagg aatgcggccg tcgcctcggc
 481 gatgttcggc aggtaggcca cgacccggtc gccggggccc accccgaggc tgcggagggc
 541 cgcagcgatc gcggcggtgc gggtccgcag ttctccccag gtccactcgg tcaacggccg
 601 gagttcggac gcgtgccgga tcgccacggc tgatgggtca cggtcgcgga agatgtgctc
 661 ggcgtagttg agggtggcgc cggggaacca gacggcgccg ggcatggcgt cggaggcgag
 721 cactgtggtg tacggggtgg cggcgcgcac ccggtagtac tcccagatcg cggaccagaa
 781 tccttcgagg tcggttaccg accagcgcca cagtgcctcg tagtccggtg cgtccacacc
 841 gcggtgctcc cgcacccagc gggtgaacgc ggtgaggttg gcgcgttctt tgcgctcctc
 901 gtcgggactc cacaggatcg gcggctgcgg cttgagtgtc atgaaacgcg acccccttcgt
 961 ggacggtgcg gatgcggtga gcgtcgggtg cctcccctaa cgctccccgg tgacggagtg
1021 ttgtgcacca catctagcac gcgggacgcg gaaaccgtat ggagaaaaca cctacaaccc
1081 cggccggacg gtgggtttcg gccacactta ggggtcgggt gcctgcttgc cgggcagggc
1141 agtcccgggg tgctgtggtg cgggcgggag ggctgtcgct tcgaggtgtg ccggcgggac
1201 actccgggcc tcagccgtac ccgcaacggg gacagttctc ctcccttccg ggctggatgg
1261 tcccttcccc cgaaatgcgg cgagatctcc cagtcagccc ggaaaacacc cgctgtgccc
1321 aggtactctt tgcttcgaac agacaggccg gacggtccac gggggaggtt tgtgggcagc
1381 ggaccacgtg cggcgaccag acgacggttg ttcctcggta tccccgctct tgtacttgtg
1441 acagcgctca cgctggtctt ggctgtcccg acggggcgcg agacgctgtg gcgcatgtgg
1501 tgtgaggcca cccaggactg gtgcctgggg gtgccggtcg actcccgcgg acagcctgcg
1561 gaggacggcg agtttctgct gctttctccg gtccaggcag cgacctgggg gaactattac
1621 gcgctcgggg attcgtactc ttcgggggac ggggcccgcg actactatcc cggcaccgcg
1681 gtgaagggcg gttgctggcg gtccgctaac gcctatccgg agctggtcgc cgaagcctac
1741 gacttcgccg gacacttgtc gttcctggcc tgcagcggcc agcgcggcta cgccatgctt
1801 gacgctatcg acgaggtcgg ctcgcagctg gactggaact cccctcacac gtcgctggtg
1861 acgatcggga tcggcggcaa cgatctgggg ttctccacgg ttttgaagac ctgcatggtg
1921 cgggtgccgc tgctggacag caaggcgtgc acggaccagg aggacgctat ccgcaagcgg
1981 atggcgaaat tcgagacgac gtttgaagag ctcatcagcg aagtgcgcac ccgcgcgccg
2041 gacgcccgga tccttgtcgt gggctacccc cggatttttc cggaggaacc gaccggcgcc
2101 tactacacgc tgaccgcgag caaccagcgg tggctcaacg aaaccattca ggagttcaac
2161 cagcagctcg ccgaggctgt cgcggtccac gacgaggaga ttgccgcgtc gggcggggtg
2221 ggcagcgtgg agttcgtgga cgtctaccac gcgttggacg gccacgagat cggctcggac
2281 gagccgtggg tgaacggggt gcagttgcgg gacctcgcca ccggggtgac tgtggaccgc
2341 agtaccttcc accccaacgc cgctgggcac cgggcggtcg gtgagcgggt catcgagcag
2401 atcgaaaccg gcccgggccg tccgctctat gccactttcg cggtggtggc gggggcgacc
2461 gtggacactc tcgcgggcga ggtggggtga cccggcttac cgtccggccc gcaggtctgc
2521 gagcactgcg gcgatctggt ccactgccca gtgcagttcg tcttcggtga tgaccagcgg
2581 cggggagagc cggatcgttg agccgtgcgt gtctttgacg agcacacccc gctgcaggag
2641 ccgttcgcac agttctcttc cggtggccag agtcgggtcg acgtcgatcc cagcccacag
2701 gccgatgctg cgggccgcga ccacgccgtt gccgaccagt tggtcgaggc gggcgcgcag
2761 cacgggggcg agggcgcgga catggtccag gtaagggccg tcgcggacga ggctcaccac
2821 ggcagtgccg accgcgcagg cgagggcgtt gccgccgaag gtgctgccgt gctggccggg
2881 gcggatcacg tcgaagactt ccgcgtcgcc taccgccgcc gccacgggca ggatgccgcc
2941 gcccagcgct ttgccgaaca ggtagatatc ggcgtcgact ccgctgtggt cgcaggcccg
```

FIGURE 33 (SEQ ID No. 40)

```
  1 vgsgpraatr rrlflgipal vlvtaltlvl avptgretlw rmwceatqdw clgvpvdsrg
 61 qpaedgefll lspvqaatwg nyyalgdsys sgdgardyyp gtavkggcwr sanaypelva
121 eaydfaghls flacsgqrgy amldaidevg sqldwnspht slvtigiggn dlgfstvlkt
181 cmvrvpllds kactdqedai rkrmakfett feelisevrt rapdarilvv gyprifpeep
241 tgayytltas nqrwlnetiq efnqqlaeav avhdeeiaas ggvgsvefvd vyhaldghei
301 gsdepwvngv qlrdlatgvt vdrstfhpna aghravgerv ieqietgpgr plyatfavva
361 gatvdtlage vg
```

FIGURE 34 (SEQ ID No. 41)

```
  1 mrttviaasa llllagcadg areetagapp gessggiree gaeastsitd vyialgdsya
 61 amggrdqplr gepfclrssg nypellhaev tdltcqgavt gdlleprtlg ertlpaqvda
121 ltedttlvtl siggndlgfg evagcireri agenaddcvd llgetigeql dqlppqldrv
181 heairdragd aqvvvtgylp lvsagdcpel gdvseadrrw aveltgqine tvreaaerhd
241 alfvlpddad ehtscappqq rwadiqgqqt dayplhptsa gheamaaavr dalglepvqp
```

FIGURE 35 (SEQ ID No. 42)

```
   1 ttctggggtg ttatggggtt gttatcggct cgtcctgggt ggatcccgcc aggtggggta
  61 ttcacggggg acttttgtgt ccaacagccg agaatgagtg ccctgagcgg tgggaatgag
 121 gtgggcgggg ctgtgtcgcc atgagggggc ggcgggctct gtggtgcccc gcgacccccg
 181 gccccggtga gcggtgaatg aaatccggct gtaatcagca tcccgtgccc accccgtcgg
 241 ggaggtcagc gcccggagtg tctacgcagt cggatcctct cggactcggc catgctgtcg
 301 gcagcatcgc gctcccgggt cttggcgtcc ctcggctgtt ctgcctgctg tccctggaag
 361 gcgaaatgat caccggggag tgatacaccg gtggtctcat cccggatgcc cacttcggcg
 421 ccatccggca attcgggcag ctccgggtgg aagtaggtgg catccgatgc gtcggtgacg
 481 ccatagtggg cgaagatctc atcctgctcg agggtgctca ggccactctc cggatcgata
 541 tcgggggcgt ccttgatggc gtccttgctg aaaccgaggt gcagcttgtg ggcttccaat
 601 ttcgcaccac ggagcgggac gaggctggaa tgacggccga agagcccgtg gtggacctca
 661 acgaaggtgg gtagtcccgt gtcatcattg aggaacacgc cctccaccgc acccagcttg
 721 tggccggagt tgtcgtaggc gctggcatcc agaagggaaa cgatctcata tttgtcggtg
 781 tgctcagaca tgatcttcct ttgctgtcgg tgtctggtac taccacggta gggctgaatg
 841 caactgttat ttttctgtta ttttaggaat tggtccatat cccacaggct ggctgtggtc
 901 aaatcgtcat caagtaatcc ctgtcacaca aaatgggtgg tgggagccct ggtcgcggtt
 961 ccgtgggagg cgccgtgccc cgcaggatcg tcggcatcgg cggatctggc cggtaccccg
1021 cggtgaataa aatcattctg taaccttcat cacggttggt tttaggtatc cgccccttc
1081 gtcctgaccc cgtccccggc gcgcgggagc ccgcgggttg cggtagacag gggagacgtg
1141 gacaccatga ggacaacggt catcgcagca agcgcattac tccttctcgc cggatgcgcg
1201 gatggggccc gggaggagac cgccggtgca ccgccgggtg agtcctccgg gggcatccgg
1261 gaggaggggg cggaggcgtc gacaagcatc accgacgtct acatcgccct cggggattcc
1321 tatgcggcga tgggcgggcg ggatcagccg ttacggggtg agccgttctg cctgcgctcg
1381 tccggtaatt acccggaact cctccacgca gaggtcaccg atctcacctg ccaggggcg
1441 gtgaccgggg atctgctcga acccaggacg ctgggggagc gcacgctgcc ggcgcaggtg
1501 gatgcgctga cggaggacac caccctggtc accctctcca tcgggggcaa tgacctcgga
1561 ttcggggagg tggcgggatg catccgggaa cggatcgccg gggagaacgc tgatgattgc
1621 gtggacctgc tgggggaaac catcggggag cagctcgatc agcttccccc gcagctggac
1681 cgcgtgcacg aggctatccg ggaccgcgcc ggggacgcgc aggttgtggt caccggttac
1741 ctgccgctcg tgtctgccgg ggactgcccc gaactggggg atgtctccga ggcggatcgt
1801 cgttgggcgg ttgagctgac cgggcagatc aacgagaccg tgcgcgaggc ggccgaacga
1861 cacgatgccc tctttgtcct gcccgacgat gccgatgagc acaccagttg tgcaccccca
1921 cagcagcgct gggcggatat ccagggccaa cagaccgatg cctatccgct gcacccgacc
1981 tccgccggcc atgaggcgat ggccgccgcc gtccgggacg cgctgggcct ggaaccggtc
2041 cagccgtagc gccgggcgcg cgcttgtcga cgaccaaccc atgccaggct gcagtcacat
2101 ccgcacatag cgcgcgcggg cgatggagta cgcaccatag aggatgagcc cgatgccgac
2161 gatgatgagc agcacactgc cgaagggttg ttccccgagg gtgcgcagag ccgagtccag
2221 acctgcggcc tgctccggat catgggccca accggcgatg acgatcaaca cccccaggat
2281 cccgaaggcg ataccacggg cgacataacc ggctgttccg gtgatgatga tcgcggtccc
2341 gacctgccct gaccccgcac ccgcctccag atcctcccgg aaatcccggg tggccccctt
2401 ccagaggttg tagacacccg cccccagtac caccagcccg gcgaccacaa ccagcaccac
2461 accccagggt tgggatagga cggtggcggt gacatcggtg gcggtctccc catcggaggt
2521 gctgccgccc cgggcgaagg tggaggtggt caccgccagg gagaagtaga ccatggccat
2581 gaccgccccc ttggcccttt ccttgaggtc ctcgcccgcc agcagctggc tcaattgcca
2641 gagtcccagg gccgccaggg cgatgacggc aacccacagg aggaactgcc cacccggagc
2701 ctccgcgatg gtggccaggg cacctgaatt cgaggcctca tcacccgaac cgccggatcc
2761 agtggcgatg cgcaccgcga tccacccgat gaggatgtgc agtatgccca ggacaatgaa
2821 accacctctg gccagggtgg tcagcgcggg gtggtcctcg gcctggtcgg cagcccgttc
2881 gatcgtccgt ttcgcggatc tggtgtcgcc cttatccata gctcccattg aaccgccttg
2941 aggggtgggc ggccactgtc agggcggatt gtgatctgaa ctgtgatgtt ccatcaaccc
```

FIGURE 36 (SEQ ID No. 43)

```
  1 mrrfrlvgfl sslvlaagaa ltgaataqaa qpaaadgyva lgdsyssgvg agsyisssgd
 61 ckrstkahpy lwaaahspst fdftacsgar tgdvlsgqlg plssgtglvs isiggndagf
121 adtmttcvlq sessclsria taeayvdstl pgkldgvysa isdkapnahv vvigyprfyk
181 lgttciglse tkrtainkas dhlntvlaqr aaahgftfgd vrttftghel csgspwlhsv
241 nwlnigesyh ptaagqsggy lpvlngaa
```

FIGURE 37 (SEQ ID No. 44)

```
   1 cccggcggcc cgtgcaggag cagcagccgg cccgcgatgt cctcgggcgt cgtcttcatc
  61 aggccgtcca tcgcgtcggc gaccggcgcc gtgtagttgg cccggacctc gtcccaggtg
 121 cccgcggcga tctggcgggt ggtgcggtgc gggccgcgcc gaggggagac gtaccagaag
 181 cccatcgtca cgttctccgg ctgcggttcg ggctcgtccg ccgctccgtc cgtcgcctcg
 241 ccgagcacct tctcggcgag gtcggcgctg gtcgccgtca ccgtgacgtc ggcgccccgg
 301 ctccagcgcg agatcagcag cgtccagccg tcgccctccg ccagcgtcgc gctgcggtcg
 361 tcgtcgcggg cgatccgcag cacgcgcgcg ccgggcggca gcagcgtggc gccggaccgt
 421 acgcggtcga tgttcgccgc gtgcgagtac ggctgctcac ccgtggcgaa acggccgagg
 481 aacagcgcgt cgacgacgtc ggacggggag tcgctgtcgt ccacgttgag ccggatcggc
 541 agggcttcgt gcgggttcac ggacatgtcg ccatgatcgg gcacccggcc gccgcgtgca
 601 cccgctttcc cgggcacgca cgacaggggc tttctcgccg tcttccgtcc gaacttgaac
 661 gagtgtcagc catttcttgg catggacact tccagtcaac gcgcgtagct gctaccacgg
 721 ttgtggcagc aatcctgcta agggaggttc catgagacgt ttccgacttg tcggcttcct
 781 gagttcgctc gtcctcgccg ccggcgccgc cctcaccggg gcagcgaccg cccaggcggc
 841 ccaacccgcc gccgccgacg gctatgtggc cctcggcgac tcctactcct ccggggtcgg
 901 agcgggcagc tacatcagct cgagcggcga ctgcaagcgc agcacgaagg cccatcccta
 961 cctgtgggcg gccgcccact cgccctccac gttcgacttc accgcctgtt ccggcgcccg
1021 tacgggtgat gttctctccg gacagctcgg cccgctcagc tccggcaccg gcctcgtctc
1081 gatcagcatc ggcggcaacg acgccggttt cgccgacacc atgacgacct gtgtgctcca
1141 gtccgagagc tcctgcctgt cgcggatcgc caccgccgag gcgtacgtcg actcgacgct
1201 gcccggcaag ctcgacggcg tctactcggc aatcagcgac aaggcgccga acgcccacgt
1261 cgtcgtcatc ggctacccgc gcttctacaa gctcggcacc acctgcatcg gcctgtccga
1321 gaccaagcgg acggcgatca acaaggcctc cgaccacctc aacaccgtcc tcgcccagcg
1381 cgccgccgcc cacggcttca ccttcggcga cgtacgcacc accttcaccg gccacgagct
1441 gtgctccggc agcccctggc tgcacagcgt caactggctg aacatcggcg agtcgtacca
1501 ccccaccgcg gccggccagt ccggtggcta cctgccggtc ctcaacggcg ccgcctgacc
1561 tcaggcggaa ggagaagaag aaggagcgga gggagacgag gagtgggagg ccccgcccga
1621 cggggtcccc gtccccgtct ccgtctccgt cccggtcccg caagtcaccg agaacgccac
1681 cgcgtcggac gtggcccgca ccggactccg cacctccacg cgcacggcac tctcgaacgc
1741 gccggtgtcg tcgtgcgtcg tcaccaccac gccgtcctgg cgcgagcgct cgccgcccga
1801 cgggaaggac agcgtccgcc accccggatc ggagaccgac ccgtccgcgg tcacccaccg
1861 gtagccgacc tccgcgggca gccgcccgac cgtgaacgtc gccgtgaacg cgggtgcccg
1921 gtcgtgcggc ggcggacagg cccccgagta gtgggtgcgc gagcccacca cggtcacctc
1981 caccgactgc gctgcggggc
```

FIGURE 38 (SEQ ID No. 45)

```
  1 mrrsritayv tslllavgca ltgaataqas paaaatgyva lgdsyssgvg agsylsssgd
 61 ckrsskaypy lwqaahspss fsfmacsgar tgdvlanqlg tlnsstglvs ltiggndagf
121 sdvmttcvlq sdsaclsrin takayvdstl pgqldsvyta istkapsahv avlgyprfyk
181 lggsclagls etkrsainda adylnsaiak raadhgftfg dvkstftghe icssstwlhs
241 ldllnigqsy hptaagqsgg ylpvmnsva
```

FIGURE 39 (SEQ ID No. 46)

```
   1 ccaccgccgg gtcggcggcg agtctcctgg cctcggtcgc ggagaggttg gccgtgtagc
  61 cgttcagcgc ggcgccgaac gtcttcttca ccgtgccgcc gtactcgttg atcaggccct
 121 tgcccttgct cgacgcggcc ttgaagccgg tgcccttctt gagcgtgacg atgtagctgc
 181 ccttgatcgc ggtgggggag ccggcggcga gcaccgtgcc ctcggccggg gtggcctggg
 241 cgggcagtgc ggtgaatccg cccacgaggg cgccggtcgc cacggcggtt atcgcggcga
 301 tccggatctt cttgctacgc agctgtgcca tacgagggag tcctcctctg ggcagcggcg
 361 cgcctgggtg gggcgcacgg ctgtgggggg tgcgcgcgtc atcacgcaca cggccctgga
 421 gcgtcgtgtt ccgccctggg ttgagtaaag cctcggccat ctacgggggt ggctcaaggg
 481 agttgagacc ctgtcatgag tctgacatga gcacgcaatc aacggggccg tgagcacccc
 541 ggggcgaccc cggaaagtgc cgagaagtct tggcatggac acttcctgtc aacacgcgta
 601 gctggtacga cggttacggc agagatcctg ctaaagggag gttccatgag acgttcccga
 661 attacggcat acgtgacctc actcctcctc gccgtcggct gcgccctcac cggggcagcg
 721 acggcgcagg cgtccccagc cgccgcggcc acgggctatg tggccctcgg cgactcgtac
 781 tcgtccggtg tcggcgccgg cagctacctc agctccagcg gcgactgcaa gcgcagttcg
 841 aaggcctatc cgtacctctg gcaggccgcg cattcaccct cgtcgttcag tttcatggct
 901 tgctcgggcg ctcgtacggg tgatgtcctg gccaatcagc tcggcaccct gaactcgtcc
 961 accggcctgg tctccctcac catcggaggc aacgacgcgg gcttctccga cgtcatgacg
1021 acctgtgtgc tccagtccga cagcgcctgc ctctcccgca tcaacacggc gaaggcgtac
1081 gtcgactcca ccctgcccgg ccaactcgac agcgtgtaca cggcgatcag cacgaaggcc
1141 ccgtcggccc atgtggccgt gctgggctac ccccgcttct acaaactggg cggctcctgc
1201 ctcgcgggcc tctcggagac caagcggtcc gccatcaacg acgcggccga ctatctgaac
1261 agcgccatcg ccaagcgcgc cgccgaccac ggcttcacct tcggcgacgt caagagcacc
1321 ttcaccggcc atgagatctg ctccagcagc acctggctgc acagtctcga cctgctgaac
1381 atcggccagt cctaccaccc gaccgcggcc ggccagtccg gcggctatct gccggtcatg
1441 aacagcgtgg cctgagctcc cacggcctga atttttaagg cctgaatttt taaggcgaag
1501 gtgaaccgga agcggaggcc ccgtccgtcg gggtctccgt cgcacaggtc accgagaacg
1561 gcacggagtt ggacgtcgtg cgcaccgggt cgcgcacctc gacggcgatc tcgttcgaga
1621 tcgttccgct cgtgtcgtac gtggtgacga acacctgctt ctgctgggtc tttccgccgc
1681 tcgccgggaa ggacagcgtc ttccagcccg gatccgggac ctcgcccttc ttggtcaccc
1741 agcggtactc cacctcgacc ggcacccggc ccaccgtgaa ggtcgccgtg aacgtgggcg
1801 cctgggcggt gggcggcggg caggcaccgg agtagtcggt gtgcacgccg gtgaccgtca
1861 ccttcacgga ctgggccggc ggggtcgtcg taccgccgcc gccaccgccg cctcccggag
1921 tggagcccga gctgtggtcg cccccgccgt cggcgttgtc gtcctcgggg gttttcgaac
```

FIGURE 40 (SEQ ID No. 47)

```
  1 mgsgpraatr rrlflgipal vlvtaltlvl avptgretlw rmwceatqdw clgvpvdsrg
 61 qpaedgefll lspvqaatwg nyyalgdsys sgdgardyyp gtavkggcwr sanaypelva
121 eaydfaghls flacsgqrgy amldaidevg sqldwnspht slvtigiggn dlgfstvlkt
181 cmvrvpllds kactdqedai rkrmakfett feelisevrt rapdarilvv gyprifpeep
241 tgayytltas nqrwlnetiq efnqqlaeav avhdeeiaas ggvgsvefvd vyhaldghei
301 gsdepwvngv qlrdlatgvt vdrstfhpna aghravgerv ieqietgpgr plyatfavva
361 gatvdtlage vg
```

FIGURE 41 (SEQ ID No. 48)

```
1       ctgcagacac ccgccccgcc ttctcccgga tcgtcatgtt cggcgactcc ctcagcgaca
61      ccggcaagat gtactccaag atgcgcggct acctgccgtc ctccccgccg tactacgagg
121     gccgcttctc gaacggcccg gtctggctgg agcagctgac gaagcagttc cccggcctga
181     cgatcgccaa cgaggccgag gggggcgcga ccgcagtcgc ctacaacaag atctcctgga
241     acccgaagta ccaggtcatt aacaacctcg actacgaggt cacccagttc ttgcagaagg
301     actcgttcaa gcccgacgac ctggtcatcc tgtgggtggg cgccaacgac tacctggcct
361     acggttggaa cacggagcag gacgccaagc gggtgcgcga cgccatctcg gacgcggcaa
421     accgcatggt cctgaacggc gcgaagcaga tcctgctgtt caacctgccc gacctgggcc
481     agaacccgtc cgcccgctcc cagaaggtcg tcgaggccgt ctcgcacgtg tccgcctacc
541     acaacaagct gctcctcaac ctcgcccggc agctcgcccc gacgggcatg gtcaagctgt
601     tcgagatcga caagcagttc gcggagatgc tgcgcgaccc ccagaacttc ggcctgagcg
661     acgtggagaa cccgtgctac gacggcggct acgtgtggaa gccgttcgcc acccggtccg
721     tctcgaccga ccggcagctg tcggccttct cgccccagga gcgcctggcg atcgctggca
781     acccgctcct ggcacaggcg gtagcttcgc cgatggcccg ccgctcggcc tcgcccctca
841     actgcgaggg caagatgttc tgggaccagg tccaccccac caccgtggtc cacgccgccc
901     tctcggagcg cgccgccacc ttcatcgaga cccagtacga gttcctcgcc cactagtcta
961     gaggatcc
```

FIGURE 42

1. L131
2. S.avermitilis
3. T.fusca
4. Consensus

```
          1                                                  50
1   (1)  --------MRLTRSLSAASVIVFALLLALLGISPAQAAG----------
2   (1)  --------MRRSRITAYVTSLLLAVGCALTGAATAQASPA----------
3   (1)  VGSGPRAATRRRLFLGIPALVLVTALTLVLAVPTGRETLWRMWCEATQDW
4   (1)          MRRSRFLA  ALILLTLA AL GAA ARAAP

               51                                            100
1   (32)  ----------------------------P-AYVALGDSYSSGNGAGSYID
2   (33)  ----------------------------AAATGYVALGDSYSSGVGAGSYLS
3   (51)  CLGVPVDSRGQPAEDGEFLLLSPVQAATWGNYYALGDSYSSGDGARDYYP
4   (51)                                A A  YVALGDSYSSG GAGSY

              101                                            150
1    (53)  SSGD---CHRSNNAYPARWAAANAP---SSFTFAACSGAVTTDVIN----
2    (57)  SSGD---CKRSSKAYPYLWQAAHSP---SSFSFMACSGARTGDVLA----
3   (101)  GTAVKGGCWRSANAYPELVAEAYDFA--GHLSFLACSGQRGYAMLDAIDE
4   (101)  SSGD   C RSTKAYPALWAAAHA    SSFSF ACSGARTYDVLA

              151                                            200
1    (93)  --NQLGALNAST--GLVSITIGGNDAGFADAMTTCVTS------SDSTCL
2    (97)  --NQLGTLNSST--GLVSLTIGGNDAGFSDVMTTCVLQ------SDSACL
3   (149)  VGSQLDWNSPHT--SLVTIGIGGNDLGFSTVLKTCMVR------VPLLDS
4   (151)     QL  LNS T   LVSITIGGNDAGFAD MTTCVL       SDSACL

              201                                            250
1   (133)  NRLATATNYINTTLLA--------RLDAVYSQIKARAPNARVVVLGYPRMY
2   (137)  SRINTAKAYVDSTLPG--------QLDSVYTAISTKAPSAHVAVLGYPRFY
3   (191)  KACTDQEDAIRKRMAKF----ETTFEELISEVRTRAPDARILVVGYPRIF
4   (201)   RIA AK YI  TLPA       RLDSVYSAI TRAP ARVVVLGYPRIY

              251                                            300
1   (176)  LASNPWYCLGLSNTKRAAINTTADTLNSVISSRATAH----------GF
2   (180)  KLGG-SCLAGLSETKRSAINDAADYLNSAIAKRAADH----------GF
3   (237)  PEEPTGAYYTLTASNQRWLNETIQEFNQQLAEAVAVHDEEIAASGGVGSV
4   (251)    SG    LGLS TKRAAINDAAD LNSVIAKRAADH          GF

              301                                            350
1   (215)  RFGDVRPTFNNHELFFGNDWLHSLTLP----------------VWESYH
2   (218)  TFGDVKSTFTGHEICSSSTWLHSLDLLN----------------IGQSYH
3   (287)  EFVDVYHALDGHEIGSDEPWVNGVQLRDLATG---------VTVDRSTFH
4   (301)  TFGDV  TF GHELCSA PWLHSLTLP                 V  SYH

              351                                        395
1   (248)  PTSTGHQSGYLPVLNANSST-----------------------
2   (252)  PTAAGQSGGYLPVMNSVA-----------------------
3   (328)  PNAAGHRAVGERVIEQIETGPGRPLYATFAVVAGATVDTLAGEVG
4   (351)  PTA GHAAGYLPVLNSI T
```

FIGURE 43

SEQ ID No 17 which is the amino acid sequence of a lipid acyltransferase from *Candida parapsilosis*;

```
MRYFAIAFLL INTISAFVLA PKKPSQDDFY TPPQGYEAQP LGSILKTRNV PNPLTNVFTP VKVQNAWQLL
VRSEDTFGNP NAIVTTIIQP FNAKKDKLVS YQTFEDSGKL DCAPSYAIQY GSDISTLTTQ GEMYYISALL
DQGYYVVTPD YEGPKSTFTV GLQSGRATLN SLRATLKSGN LTGVSSDAET LLWGYSGGSL ASGWAAAIQK
EYAPELSKNL LGAALGGFVT NITATAEAVD SGPFAGIISN ALAGIGNEYP DFKNYLLKKV SPLLSITYRL
GNTHCLLDGG IAYFGKSFFS
RIIRYFPDGW DLVNQEPIKT ILQDNGLVYQ PKDLTPQIPL FIYHGTLDAI VPIVNSRKTF QQWCDWGLKS
GEYNEDLTNG HITESIVGAP AALTWIINRF NGQPPVDGCQ HNVRASNLEY PGTPQSIKNY FEAALHAILG
FDLGPDVKRD KVTLGGLLKL ERFAF
```

FIGURE 44

SEQ ID No 18 which is the amino acid sequence of a lipid acyltransferase from *Candida parapsilosis*;

```
MRYFAIAFLL INTISAFVLA PKKPSQDDFY TPPQGYEAQP LGSILKTRNV PNPLTNVFTP VKVQNAWQLL
VRSEDTFGNP NAIVTTIIQP FNAKKDKLVS YQTFEDSGKL DCAPSYAIQY GSDISTLTTQ GEMYYISALL
DQGYYVVTPD YEGPKSTFTV GLQSGRATLN SLRATLKSGN LTGVSSDAET LLWGYSGGSL ASGWAAAIQK
EYAPELSKNL LGAALGGFVT NITATAEAVD SGPFAGIISN ALAGIGNEYP DFKNYLLKKV SPLLSITYRL
GNTHCLLDGG IAYFGKSFFS RIIRYFPDGW DLVNQEPIKT ILQDNGLVYQ PKDLTPQIPL FIYHGTLDAI
VPIVNSRKTF QQWCDWGLKS GEYNEDLTNG HITESIVGAP AALTWIINRF NGQPPVDGCQ HNVRASNLEY
PGTPQSIKNY FEAALHAILG FDLGPDVKRD KVTLGGLLKL ERFAFHHHHH H
```

FIGURE 45

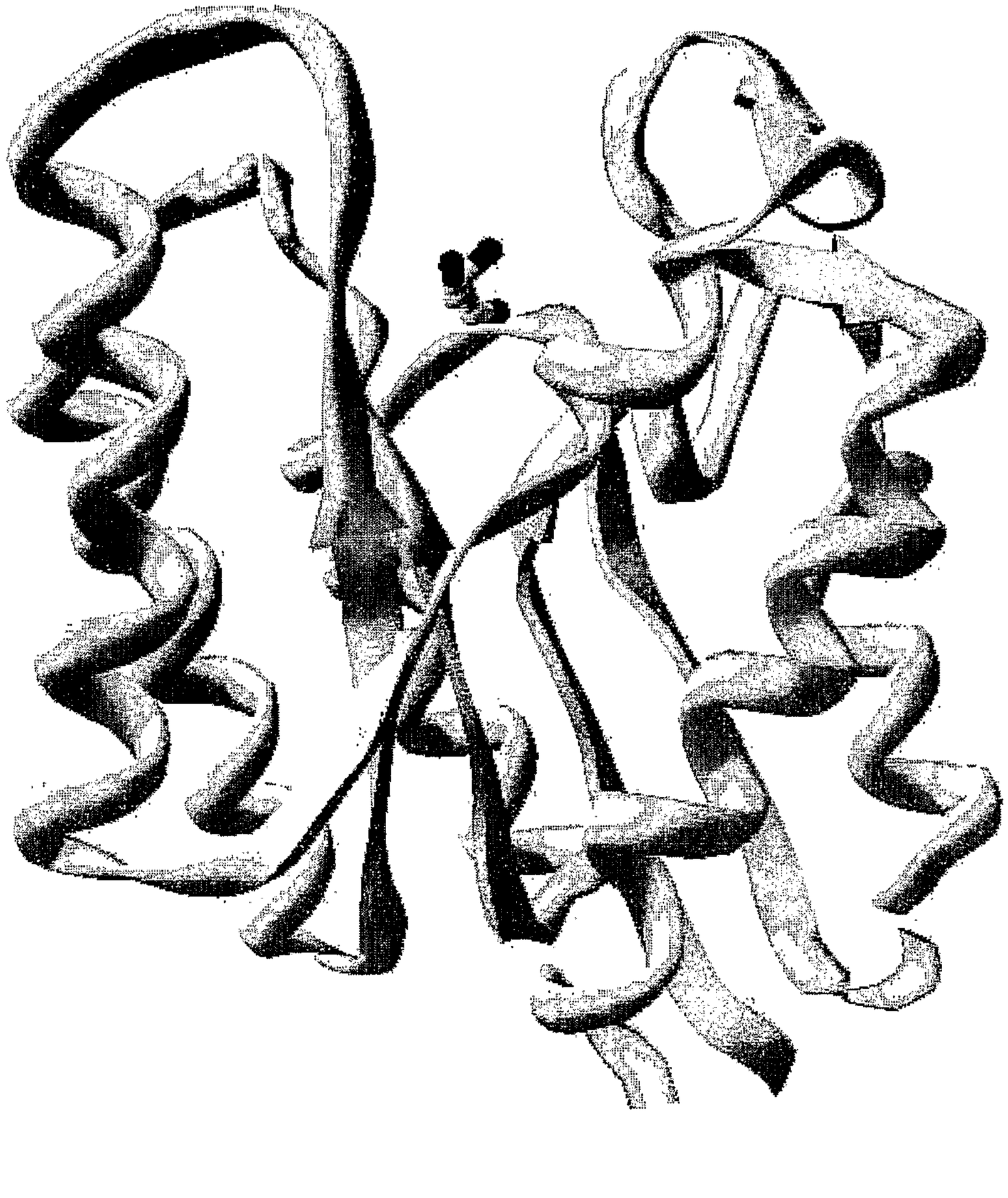


FIGURE 46

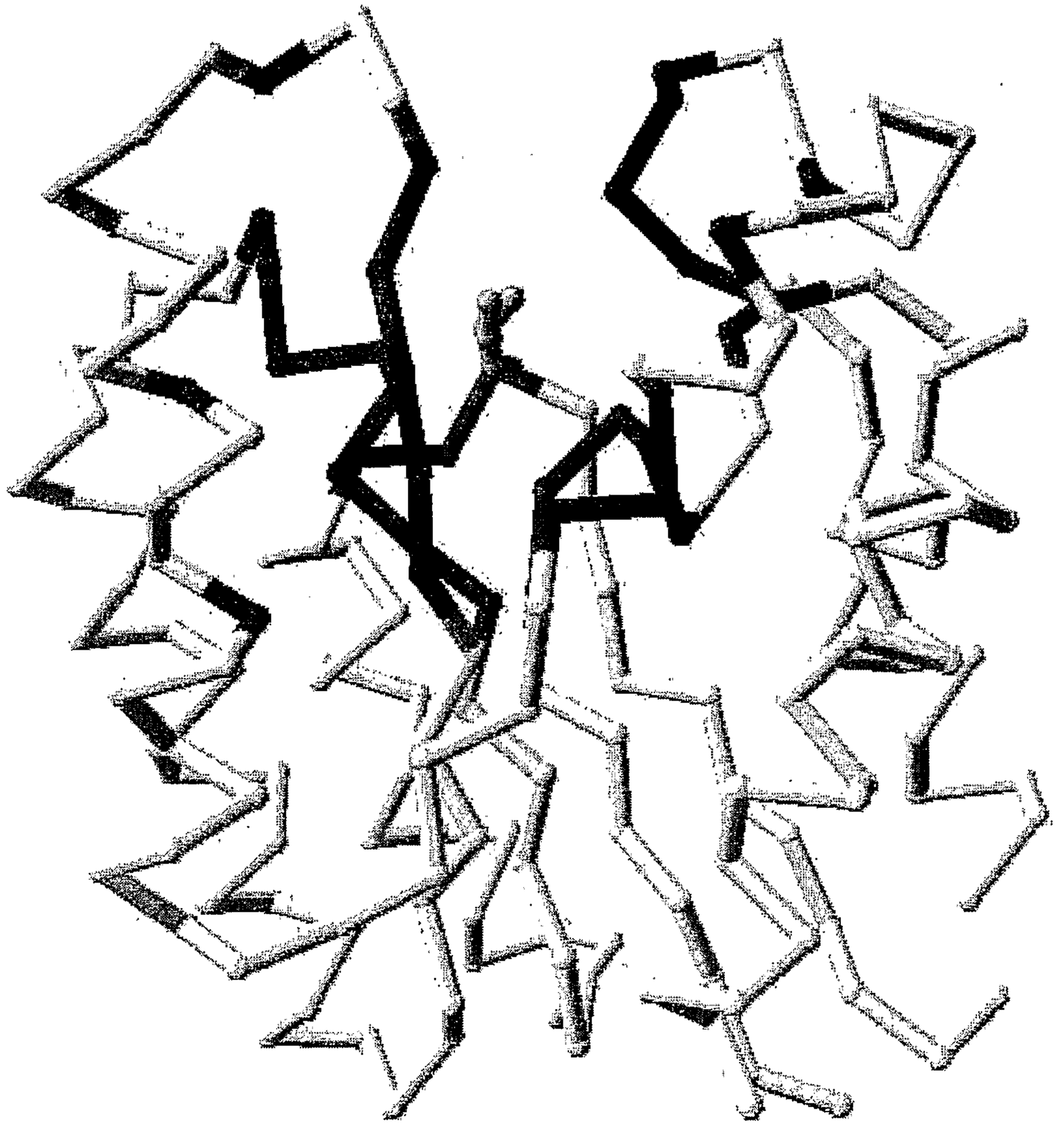
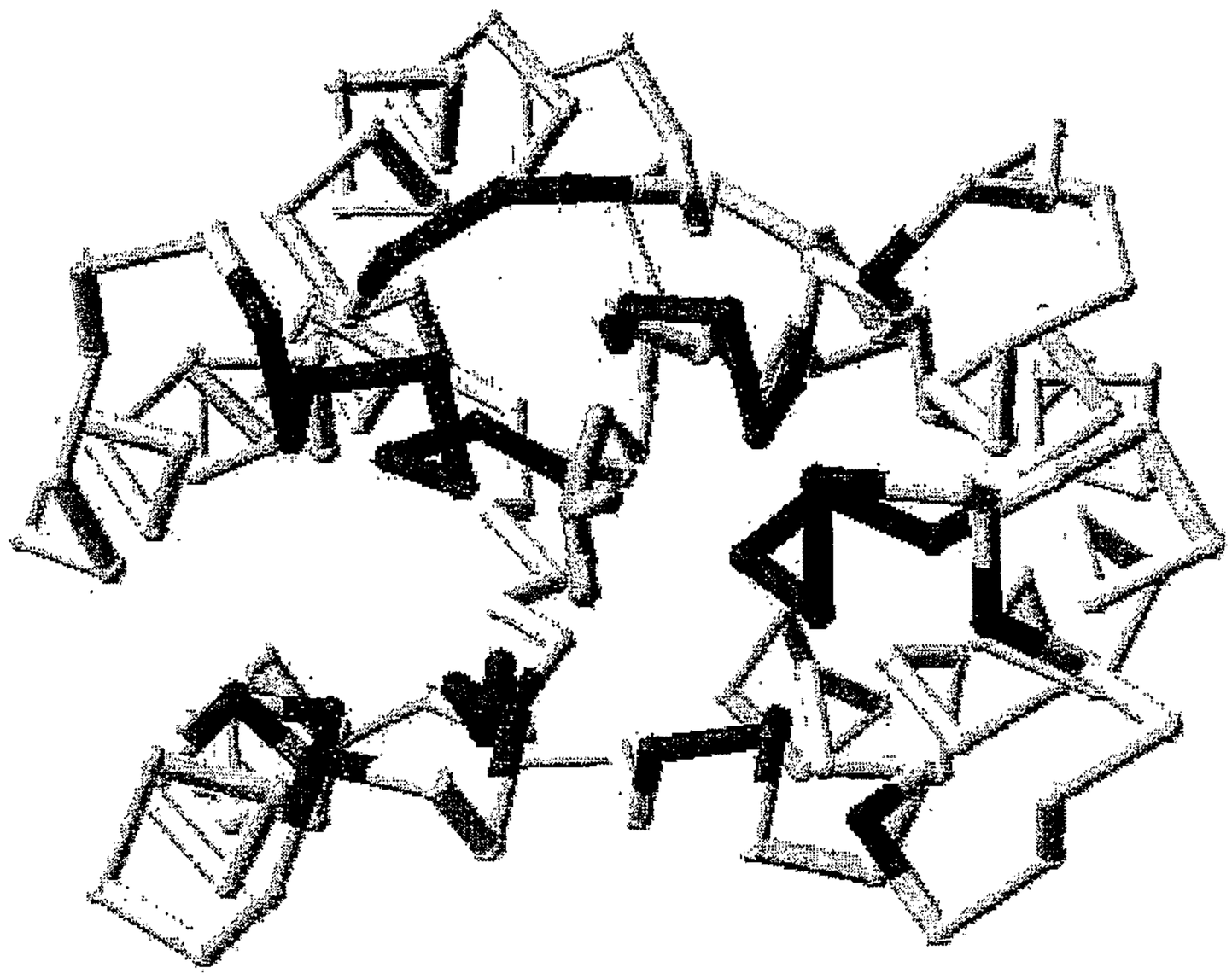


FIGURE 47



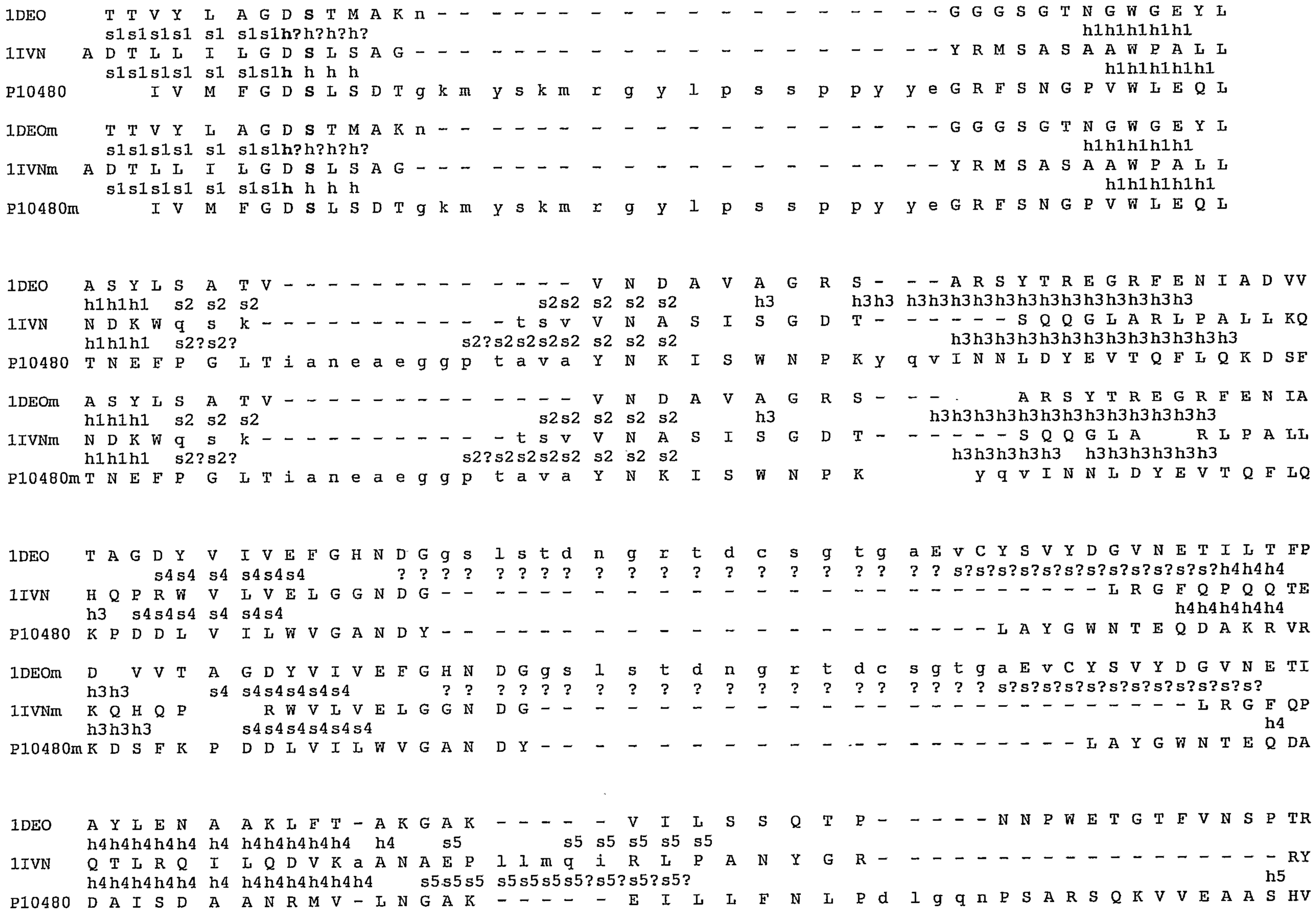
FIGRURE 48

```
1DEO     T T V Y  L  A G D S T M A K n - - - - - - - - - - - - - - - - - - - - - - G G G S G T N G W G E Y L
         s1s1s1s1 s1 s1s1h?h?h?h?                                              h1h1h1h1h1
1IVN   A D T L L  I  L G D S L S A G - - - - - - - - - - - - - - - - - - - - - - - Y R M S A S A A W P A L L
         s1s1s1s1 s1 s1s1h h h h                                                 h1h1h1h1h1
P10480       I V  M  F G D S L S D T g k m  y s k m  r  g  y  l p s s p p y  y e G R F S N G P V W L E Q L

1DEO   A S Y L S  A  T V - - - - - - - - - - - - - - -  V  N  D  A V A G R S -  - - A R S Y T R E G R F E N I A D VV
       h1h1h1  s2 s2 s2                          s2s2 s2 s2 s2     h3    h3h3 h3h3h3h3h3h3h3h3h3h3h3h3h3
1IVN   N D K W q  s  k - - - - - - - - - - - - t s v  V  N  A  S I S G D T -  - - - - - S Q Q G L A R L P A L L KQ
       h1h1h1  s2?s2?                      s2?s2s2s2s2 s2 s2 s2                     h3h3h3h3h3h3h3h3h3h3h3h3
P10480 T N E F P  G  L T i a n e a e g g p  t a v a  Y  N  K  I S W N P K y  q v I N N L D Y E V T Q F L Q K D SF

1DEO   T A G D Y  V  I V E F G H N D G g s  l s t d  n  g  r  t d c s g t g  a E v C Y S V Y D G V N E T I L T FP
             s4s4 s4 s4s4s4        ? ? ? ?  ? ? ? ?  ?  ?  ?  ? ? ? ? ? ? ?  ? ? s?s?s?s?s?s?s?s?s?s?s?s?h4h4h4
1IVN   H Q P R W  V  L V E L G G N D G - - - - - - - - - - - - - - - - - - - - - - - - - - - L R G F Q P Q Q TE
       h3  s4s4s4 s4 s4s4                                                                  h4h4h4h4h4
P10480 K P D D L  V  I L W V G A N D Y - - - - - - - - - - - - - - - - - - - - - - - - L A Y G W N T E Q D A K R VR

1DEO   A Y L E N  A  A K L F T - A K G A K  - - - -  -  V  I  L S S Q T P -  - - - - N N P W E T G T F V N S P TR
       h4h4h4h4h4 h4 h4h4h4h4h4  h4   s5          s5 s5 s5 s5 s5
1IVN   Q T L R Q  I  L Q D V K a A N A E P  l l m q  i  R  L  P A N Y G R -  - - - - - - - - - - - - - - - - - RY
       h4h4h4h4h4 h4 h4h4h4h4h4h4  s5s5s5 s5s5s5s5?s5?s5?s5?                                            h5
P10480 D A I S D  A  A N R M V - L N G A K  - - - -  -  E  I  L L F N L P d  l g q n P S A R S Q K V V E A A S HV

1DEO   F V E Y A  E  L A A E V A - - - - - - - - - - - - - - - G V E Y V  D H W S Y V D S I Y E T L G N A t vn
       h5h5h5h5h5 h5 h5h5h5h5h5h5                          s6s6s6s6?h6h6h6h6h6h6h6h6h6h6h6    h h h h
1IVN   N E A F S  A  I Y P K L A k e - - - - - - - - - - - - f D V P L L  P F F M E E V Y L K P Q W - - - - - --
       h5h5h5h5h5 h5 h5h5h5h5h5h5h5h5                  h5    s6s6s6s6?  h6h6h6h6h6     s
P10480 S A Y H N  Q  L L L N L A r q l a p  t g m v  k  l  f  e i D K Q F A  E M L R D P Q N F G L S D Q R N a cy

1DEO   - - - - -  -  - - - - - - - - - - -  - - - -  -  -  -  - - - - - - -  - - - - - - - - - - - - - - - - - --

1IVN   - - - - -  -  - - - - - - - - - - -  - - - -  -  -  -  - - - - - - -  - - - - - - - - - - - - - - - - - --

P10480 g g s y v  w  k p f a s r s a s t d  s q l s  a  f  n  p q e r l a i  a g n p l l a q a v a s p m a a r sa

1DEO   - - - - -  -  s y F P I D H T H T S  P A G A  E  V  V  A E A F L K A  V V C T G T S L K S V L T T T S F EG
                     h             s?s?s?  h7h7h7h7 h7 h7 h7 h7h7h7h7h7h7h7 h7h7h7h7    h?h?h?
1IVN   - - - - -  -  - - M Q D D G I H P N  R D A Q  P  F  I  A D W M A K Q  L Q P L V N H D S L E
                         s s         s s s    h7h7h7 h7 h7 h7 h7h7h7h7h7h7h7 h7h7h7h7
P10480 s t l n c  e  g k M F W D Q V H P T  T V V H  A  A  L  S E P A A T F  I E S Q Y E F L A H -
```

FIGURE 49

```
1DEOm    T T V Y  L  A G D S T M A K n - -  - - - - - - - - - -  -  -  -  -  - - - - G G G S G T N G W G E Y L
           s1s1s1s1 s1 s1s1h?h?h?h?                                                  h1h1h1h1h1
1IVNm   A D T L L  I  L G D S L S A G - - -  - - - - - - - - - -  -  -  -  -  - - - - Y R M S A S A A W P A L L
           s1s1s1s1 s1 s1s1h h h h                                                      h1h1h1h1h1
P10480m      I V  M  F G D S L S D T g k m  y s k m r g y l  p  s  s  p  p y y e G R F S N G P V W L E Q L

1DEOm   A S Y L S  A  T V - - - - - - - - - - -  - - - - V N D A  V  A  G  R  S - - -     A R S Y T R E G R F E N IA
        h1h1h1  s2 s2 s2                   s2s2s2s2s2     h3          h3h3h3h3h3h3h3h3h3h3h3h3h3
1IVNm   N D K W q  s  k - - - - - - - - - - -  - t s v V N A S  I  S  G  D  T - - - - - - S Q Q G L A    R L P A LL
        h1h1h1  s2?s2?               s2?s2s2s2s2s2s2s2                      h3h3h3h3h3  h3h3h3h3h3h3
P10480m T N E F P  G  L T i a n e a e g g p  t a v a Y N K I  S  W  N  P  K        y q v I N N L D Y E V T Q F LQ

1DEOm   D  V V T  A  G D Y V I V E F G H N  D G g s l s t d  n  g  r  t  d c s g t g a E v C Y S V Y D G V N E TI
        h3h3      s4 s4s4s4s4s4         ? ?  ? ? ? ? ? ? ? ?  ?  ?  ?  ?  ? ? ? ? ? ? s?s?s?s?s?s?s?s?s?s?s?s?
1IVNm   K Q H Q P     R W V L V E L G G N  D G - - - - - - -  -  -  -  -  - - - - - - - - - - - - - - - L R G F QP
        h3h3h3        s4s4s4s4s4s4                                                                         h4
P10480m K D S F K  P  D D L V I L W V G A N  D Y - - - - - - -  -  -  -  -  - - - - - - - - - - - - L A Y G W N T E Q DA

1DEOm   L T F P A  Y  L E N A A K L F T A K  G A K V I L S S  Q  T  P  N  N P W E T G T F V N S P T R
          h4h4h4h4 h4 h4h4h4h4h4h4h4h4h4h4     s5s5s5s5s5s5
1IVNm   Q Q T E Q  T  L R Q I L Q D V K a A  N A E P l l m q  i  R  L  P  A N Y G R - - - - - - - - - - R Y N E A
        h4h4h4h4h4 h4 h4h4h4h4h4h4h4h4h4h4    s5s5s5s5s5s5s5?s5?s5?s5?                              h5h5h5h5h5h5
P10480m K R V R D  A  I S D A A N R M V L N  G A K E I L L F  N  L  P  d  l g q n P S A R S Q K V V E A A S H V SA

1DEOm     F V E Y  A  E L A A E V A - - - -  - - - - - - - - G  V  E  Y  V  D H W S Y V D S I Y E T L G N A t v n --
          h5h5h5h5 h5 h5h5h5h5h5h5h5                         s6 s6 s6 s6?h6h6h6h6h6h6h6h6h6h6   h h h h
1IVNm     F S A I  Y  P K L A k e - - - - -  - - - - - - f D  V  P  L  L  P F F M E E V Y        L K P Q W - - --
        h5h5h5h5h5 h5 h5h5h5                 h5  s6s6s6s6?   h6 h6 h6 h6h6      s
P10480m Y H N Q L  L  L N L A r q l a p t g  m v k l f e i D  K  Q  F  A  E M L R D P Q N F G L S D Q R N a c y gg

1DEOm   - - - - -  -  - - - - - - - - - - -  - - - - - - - -  -  -  -  -  - - - - - - - - - - - - - - - - - - --

1IVNm   - - - - -  -  - - - - - - - - - - -  - - - - - - - -  -  -  -  -  - - - - - - - - - - - - - - - - - - --

P10480m s y v w k  p  f a s r s a s t d s q  l s a f n p q e  r  l  a  i  a g n p l l a q a v a s p m a a r s a st

1DEOm   - - - - s  y  F P I D H T H T S P A  G A E V V A E A  F  L  K  A  V V C T G T S L K S V L T T T S F E G TC
                h                s?s?s?h7h7  h7h7h7h7h7h7h7h7 h7 h7 h7 h7 h7h7h7h7    h?h?h?
1IVNm   - - - - -  -  M Q D D G I H P N R D  A Q P F I A D W  M  A  K  Q  L Q P L V N H D S L E
                   s  s s   h7h7h7h7h7h7h7h7 h7h7h7h7h7h7h7h7 h7 h7 h7 h7 h7
P10480m l n c e g  k  M F W D Q V H P T T V  V H A A L S E P  A  A  T  F  I E S Q Y E F L A H -
```

FIGURE 50


1DEO T T V Y L A G D S T M A K n - G G G S G T N G W G E Y L
s1s1s1s1 s1 s1s1h?h?h?h? h1h1h1h1h1
1IVN A D T L L I L G D S L S A G - Y R M S A S A A W P A L L
s1s1s1s1 s1 s1s1h h h h h1h1h1h1h1
P10480 I V M F G D S L S D T g k m y s k m r g y l p s s p p y y e G R F S N G P V W L E Q L
1DEOm T T V Y L A G D S T M A K n - G G G S G T N G W G E Y L
s1s1s1s1 s1 s1s1h?h?h?h? h1h1h1h1h1
1IVNm A D T L L I L G D S L S A G - Y R M S A S A A W P A L L
s1s1s1s1 s1 s1s1h h h h h1h1h1h1h1
P10480m I V M F G D S L S D T g k m y s k m r g y l p s s p p y y e G R F S N G P V W L E Q L
1DEO A S Y L S A T V - - - - - - - - - - - - - - V N D A V A G R S - - - A R S Y T R E G R F E N I A D VV
h1h1h1 s2 s2 s2 s2s2 s2 s2 s2 h3 h3h3 h3h3h3h3h3h3h3h3h3h3h3h3h3h3
1IVN N D K W q s k - - - - - - - - - - - - - t s v V N A S I S G D T - - - - - - - S Q Q G L A R L P A L L KQ
h1h1h1 s2?s2? s2?s2s2s2s2 s2 s2 s2 h3h3h3h3h3h3h3h3h3h3h3h3h3h3
P10480 T N E F P G L T i a n e a e g g p t a v a Y N K I S W N P K y q v I N N L D Y E V T Q F L Q K D SF
1DEOm A S Y L S A T V - - - - - - - - - - - - - - V N D A V A G R S - - - A R S Y T R E G R F E N IA
h1h1h1 s2 s2 s2 s2s2 s2 s2 s2 h3 h3h3h3h3h3h3h3h3h3h3h3h3h3h3
1IVNm N D K W q s k - - - - - - - - - - - - - t s v V N A S I S G D T - - - - - - - S Q Q G L A R L P A LL
h1h1h1 s2?s2? s2?s2s2s2s2 s2 s2 s2 h3h3h3h3h3 h3h3h3h3h3h3
P10480m T N E F P G L T i a n e a e g g p t a v a Y N K I S W N P K y q v I N N L D Y E V T Q F LQ
1DEO T A G D Y V I V E F G H N D G g s l s t d n g r t d c s g t g a E v C Y S V Y D G V N E T I L T FP
s4s4 s4 s4s4s4 ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? s?s?s?s?s?s?s?s?s?s?s?h4h4h4
1IVN H Q P R W V L V E L G G N D G - L R G F Q P Q Q TE
h3 s4s4s4 s4 s4s4 h4h4h4h4h4
P10480 K P D D L V I L W V G A N D Y - L A Y G W N T E Q D A K R VR
1DEOm D V V T A G D Y V I V E F G H N D G g s l s t d n g r t d c s g t g a E v C Y S V Y D G V N E TI
h3h3 s4 s4s4s4s4s4 ? s?s?s?s?s?s?s?s?s?s?s?
1IVNm K Q H Q P R W V L V E L G G N D G - L R G F QP
h3h3h3 s4s4s4s4s4s4 h4
P10480m K D S F K P D D L V I L W V G A N D Y - L A Y G W N T E Q DA
1DEO A Y L E N A A K L F T - A K G A K - - - - - - V I L S S Q T P - - - - - N N P W E T G T F V N S P TR
h4h4h4h4h4h4 h4 h4h4h4h4h4h4 h4 s5 s5 s5 s5 s5 s5
1IVN Q T L R Q I L Q D V K a A N A E P l l m q i R L P A N Y G R - RY
h4h4h4h4h4h4 h4 h4h4h4h4h4h4h4 s5s5s5 s5s5s5s5?s5?s5?s5? h5
P10480 D A I S D A A N R M V - L N G A K - - - - - - E I L L F N L P d l g q n P S A R S Q K V V E A A S HV

```
1DEOm   L T F P A  Y  L E N A A K L F T A K  G A K V  I  L  S  S  Q  T  P  N  N P  W E T G T F V N S P T R
          h4h4h4h4 h4 h4h4h4h4h4h4h4h4h4h4h4    s5s5s5 s5 s5 s5
1IVNm   Q Q T E Q  T  L R Q I L Q D V K a A  N A E P  l  l  m  q  i  R  L  P  A N  Y G R - - - - - - - - - - R Y N E A
        h4h4h4h4h4 h4 h4h4h4h4h4h4h4h4h4h4h4    s5s5s5 s5 s5 s5 s5?s5?s5?s5?                        h5h5h5h5h5h5h5
P10480mK R V R D  A  I S D A A N R M V L N  G A K E  I  L  L  F  N  L  P  d  l g  q n P S A R S Q K V V E A A S H V SA

1DEO    F V E Y A  E  L A A E V A - - - - -  - - - - - - - - - - - G  V  E  Y V  D H W S Y V D S I Y E T L G N A t vn
        h5h5h5h5h5 h5 h5h5h5h5h5h5h5                                  s6 s6 s6s6?h6h6h6h6h6h6h6h6h6h6h6    h h h h
1IVN    N E A F S  A  I Y P K L A k e - - -  - - - - - - - - - - f  D  V  P  L L  P F F M E E V Y L K P Q W - - - - --
        h5h5h5h5h5 h5 h5h5h5h5h5h5h5h5h5                    h5        s6 s6 s6s6?  h6h6h6h6h6     s
P10480  S A Y H N  Q  L L L N L A r q l a p  t g m v  k  l  f  e  i  D  K  Q  F A  E M L R D P Q N F G L S D Q R N a cy

1DEOm     F V E Y  A  E L A A E V A - - - -  - - - - - - - G  V  E  Y  V  D H  W S Y V D S I Y E T L G N A t v n --
          h5h5h5h5 h5 h5h5h5h5h5h5h5h5                       s6 s6 s6 s6?h6h6 h6h6h6h6h6h6h6h6h6    h h h h
1IVNm     F S A I  Y  P K L A k e - - - -  - - - - - - f  D  V  P  L  L  P F  F M E E V Y        L K P Q W - - --
        h5h5h5h5h5 h5 h5h5h5                  h5    s6 s6 s6 s6?   h6 h6 h6 h6h6      s
P10480mY H N Q L  L  L N L A r q l a p t g  m v k l  f  e  i  D  K  Q  F  A  E M  L R D P Q N F G L S D Q R N a c y gg

1DEO    - - - - -  -  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - --
1IVN    - - - - -  -  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - --
P10480  g g s y v  w  k p f a s r s a s t d  s q l s  a  f  n  p  q  e  r  l  a i  a g n p l l a q a v a s p m a a r sa
1DEOm   - - - - -  -  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - --
1IVNm   - - - - -  -  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - --
P10480ms y v w k  p  f a s r s a s t d s q  l s a f  n  p  q  e  r  l  a  i  a g  n p l l a q a v a s p m a a r s a st

1DEO    - - - - - -  s y F P I D H T H T S  P A G A  E  V  V  A  E  A  F  L  K A  V V C T G T S L K S V L T T T S F EG
              h                  s?s?s? h7h7h7h7 h7 h7 h7 h7 h7 h7 h7 h7 h7h7 h7h7h7h7    h?h?h?
1IVN    - - - - - -  - - M Q D D G I H P N  R D A Q  P  F  I  A  D  W  M  A  K Q  L Q P L V N H D S L E
                     s s        s s s    h7h7h7 h7 h7 h7 h7 h7 h7 h7 h7 h7h7 h7h7h7h7
P10480  s t l n c  e  g k M F W D Q V H P T  T V V H  A  A  L  S  E  P  A  A  T F  I E S Q Y E F L A H -

1DEOm   - - - - s  y  F P I D H T H T S P A  G A E V  V  A  E  A  F  L  K  A  V V  C T G T S L K S V L T T T S F E G TC
                h                  s?s?s?h7h7 h7h7h7h7h7 h7 h7 h7 h7 h7 h7 h7 h7 h7h7 h7h7    h?h?h?
1IVNm   - - - - -  -  M Q D D G I H P N R D  A Q P F  I  A  D  W  M  A  K  Q  L Q  P L V N H D S L E
                s  s s    h7h7h7h7h7h7h7h7h7 h7h7h7h7h7 h7 h7 h7 h7 h7 h7 h7 h7 h7
P10480m l n c e g  k  M F W D Q V H P T T V  V H A A  L  S  E  P  A  A  T  F  I E  S Q Y E F L A H -
```

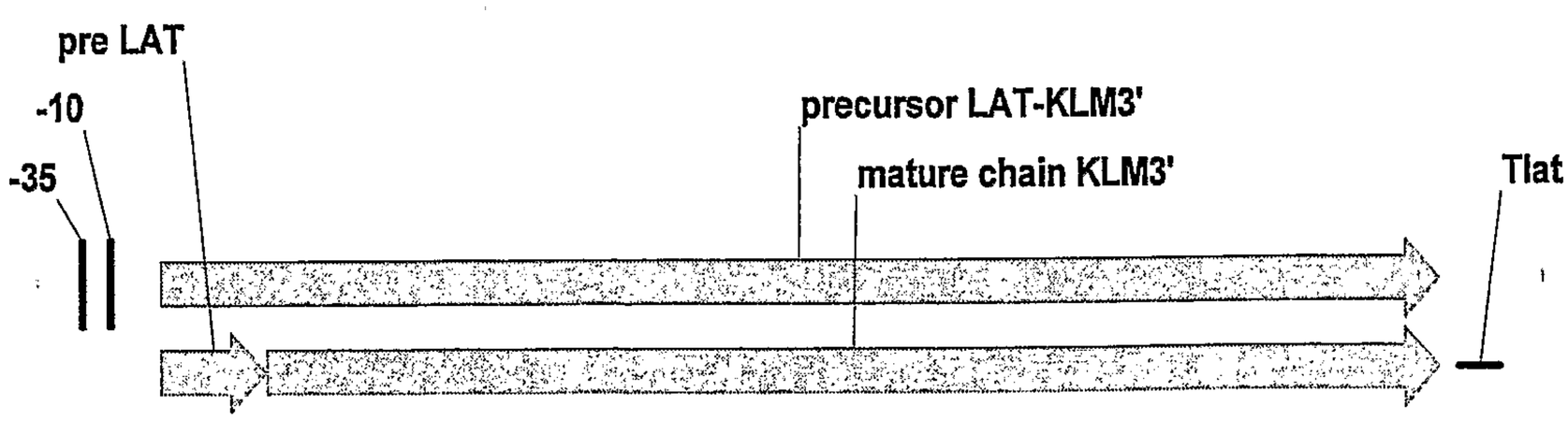
FIGURE 51

```
                                10        20        30        40        50
60

....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A          4 LLILGDSLSAG--------------------YRMSASAAWPALLNDKWqsk---
------- 34
P10480         28
IVMFGDSLSDTgkmyskmrgylpssppyyeGRFSNGPVWLEQLTNEFPGLTianeaeggp 87

                                70        80        90       100       110
120

....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A         35 -tsvVNASISGDT---------------------------
SQQGLARLPALLKQHQPRW 65
P10480         88 tavaYNKISWNPKyq----------------------
vINNLDYEVTQFLQKDSFKPDDL 125

                               130       140       150       160       170
180

....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A         66 VLVELGGNDG-----------------------------------
LRGFQPQQTEQT 87
P10480        126 VILWVGANDY-------------------------------LA--
YGWNTEQDAKRVRDA 152

                               190       200       210       220       230
240

....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A         88 LRQILQDVKaANAEPllmqiRLPANYGR-------------------------
------- 115
P10480        153 ISDAANRMV-LNGAK-----EILLFNLPdlg----------------------
----qnP 180

                               250       260       270       280       290
300

....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A        116 ------------RYNEAFSAIYPKLAke----------------------
fDVPLLPFFME 142
P10480        181 SARSQKVVEAASHVSAYHNQLLLNLArqlaptg----------
mvklfeiDKQFAEMLRD 230

                               310       320       330       340       350
360

....*....|....*....|....*....|....*....|....*....|....*....|....*....|
1IVN_A        143 EVYLKPQW-----------------------------------------
------- 150
P10480        231
PQNFGLSDQRNacyggsyvwkpfasrsastdsqlsafnpqerlaiagnpllaqavaspma 290

                               370       380       390       400
                    ....*....|....*....|....*....|....*....|....*....|
1IVN_A        151 ------------MQDDGI--------HPNRDAQPFIADWM 170
P10480        291 arsastlncegkMFWDQV--------HPTTVVHAALSEPA 322
```

FIGURE 52

```
                 1                                                  50
   P10480    (1) MKKWFVCLLGLVALTVQAADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLP
   A. sal    (1) ------------------ADTRPAFSRIVMFGDSLSDTGKMYSKMRGYLP
   A. hyd    (1) ------------------ADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLP
Consensus    (1)                   AD*RPAFSRIVMFGDSLSDTGKMYSKMRGYLP
                51                                                 100
   P10480   (51) SSPPYYEGRFSNGPVWLEQLTNEFPGLTIANEAEGGPTAVAYNKISWNPK
   A. sal   (33) SSPPYYEGRFSNGPVWLEQLTKQFPGLTIANEAEGGATAVAYNKISWNPK
   A. hyd   (33) SSPPYYEGRFSNGPVWLEQLTKQFPGLTIANEAEGGATAVAYNKISWNPK
Consensus   (51) SSPPYYEGRFSNGPVWLEQLT**FPGLTIANEAEGG*TAVAYNKISWNPK
               101                                                 150
   P10480  (101) YQVINNLDYEVTQFLQKDSFKPDDLVILWVGANDYLAYGWNTEQDAKRVR
   A. sal   (83) YQVINNLDYEVTQFLQKDSFKPDDLVILWVGANDYLAYGWNTEQDAKRVR
   A. hyd   (83) YQVINNLDYEVTQFLQKDSFKPDDLVILWVGANDYLAYGWNTEQDAKRVR
Consensus  (101) YQVINNLDYEVTQFLQKDSFKPDDLVILWVGANDYLAYGWNTEQDAKRVR
               151                                                 200
   P10480  (151) DAISDAANRMVLNGAKEILLFNLPDLGQNPSARSQKVVEAASHVSAYHNQ
   A. sal  (133) DAISDAANRMVLNGAKQILLFNLPDLGQNPSARSQKVVEAVSHVSAYHNK
   A. hyd  (133) DAISDAANRMVLNGAKQILLFNLPDLGQNPSARSQKVVEAVSHVSAYHNQ
Consensus  (151) DAISDAANRMVLNGAK*ILLFNLPDLGQNPSARSQKVVEA*SHVSAYHN*
               201                                                 250
   P10480  (201) LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDQRNACYGGSYVW
   A. sal  (183) LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDVENPCYDGGYVW
   A. hyd  (183) LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDVENPCYDGGYVW
Consensus  (201) LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSD**N*CY*G*YVW
               251                                                 300
   P10480  (251) KPFASRSASTDSQLSAFNPQERLAIAGNPLLAQAVASPMAARSASTLNCE
   A. sal  (233) KPFATRSVSTDRQLSAFSPQERLAIAGNPLLAQAVASPMARRSASPLNCE
   A. hyd  (233) KPFATRSVSTDRQLSAFSPQERLAIAGNPLLAQAVASPMARRSASPLNCE
Consensus  (251) KPFA*RS*STD*QLSAF*PQERLAIAGNPLLAQAVASPMA*RSAS*LNCE
               301                                         336
   P10480  (301) GKMFWDQVHPTTVVHAALSEPAATFIESQYEFLAH-
   A. sal  (283) GKMFWDQVHPTTVVHAALSERAATFIETQYEFLAHG
   A. hyd  (283) GKMFWDQVHPTTVVHAALSERAATFIANQYEFLAH-
Consensus  (301) GKMFWDQVHPTTVVHAALSE*AATFI**QYEFLAH*
```

FIGURE 53



Gene construct for KLM3' expression

1225 bp

FIGURE 54

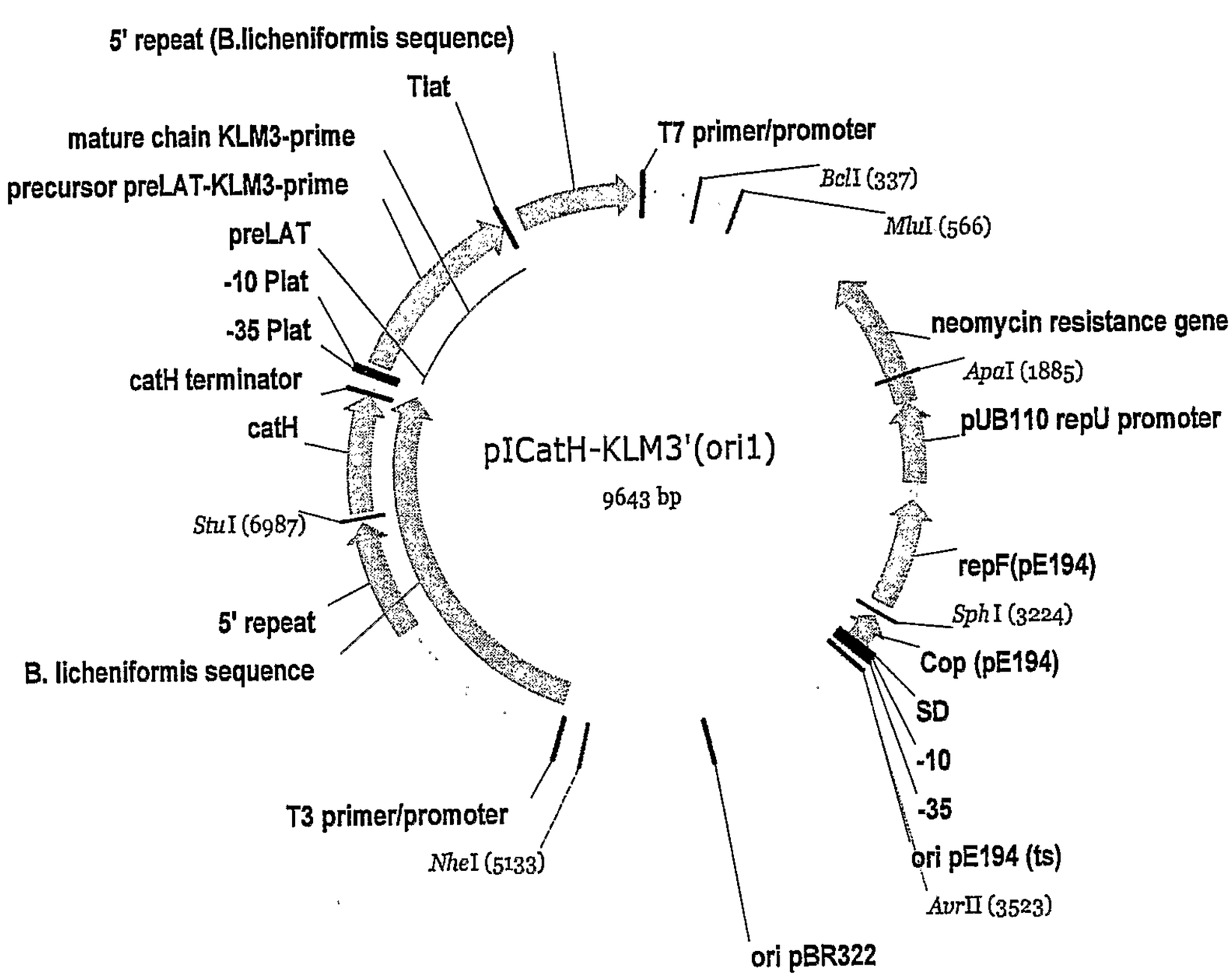

5' repeat (B.licheniformis sequence)
Tlat
mature chain KLM3-prime
precursor preLAT-KLM3-prime
preLAT
-10 Plat
-35 Plat
catH terminator
catH
StuI (6987)
5' repeat
B. licheniformis sequence
T7 primer/promoter
BclI (337)
MluI (566)
neomycin resistance gene
ApaI (1885)
pUB110 repU promoter
repF(pE194)
SphI (3224)
Cop (pE194)
SD
-10
-35
ori pE194 (ts)
AvrII (3523)
pICatH-KLM3'(ori1)
9643 bp
T3 primer/promoter
NheI (5133)
ori pBR322

FIGURE 55

```
                                                    -35
  1  GCTTTTCTTT TGGAAGAAAA TATAGGGAAA ATGGTACTTG TTAAAAATTC GGAATATTTA
     CGAAAAGAAA ACCTTCTTTT ATATCCCTTT TACCATGAAC AATTTTTAAG CCTTATAAAT
       -10                                      M  K  Q  Q  K  R  L  .
 61  TACAATATCA TATGTTTCAC ATTGAAAGGG GAGGAGAATC ATGAAACAAC AAAAACGGCT
     ATGTTATAGT ATACAAAGTG TAACTTTCCC CTCCTCTTAG TACTTTGTTG TTTTTGCCGA
     .  Y  A  R     L  L  T  L     L  F  A     L  I  F     L  L  P  H     S  A  A  .
121  TTACGCCCGA TTGCTGACGC TGTTATTTGC GCTCATCTTC TTGCTGCCTC ATTCTGCAGC
     AATGCGGGCT AACGACTGCG ACAATAAACG CGAGTAGAAG AACGACGGAG TAAGACGTCG
     .  S  A  A     D  T  R  P     A  F  S     R  I  V     M  F  G  D     S  L  S  .
181  TTCAGCAGCA GATACAAGAC CGGCGTTTAG CCGGATCGTC ATGTTTGGAG ATAGCCTGAG
     AAGTCGTCGT CTATGTTCTG GCCGCAAATC GGCCTAGCAG TACAAACCTC TATCGGACTC
     .  D  T  G     K  M  Y  S     K  M  R     G  Y  L     P  S  S  P     P  Y  Y  .
241  CGATACGGGC AAAATGTATA GCAAAATGAG AGGCTATCTT CCGTCAAGCC CGCCGTATTA
     GCTATGCCCG TTTTACATAT CGTTTTACTC TCCGATAGAA GGCAGTTCGG GCGGCATAAT
     .  E  G  R     F  S  N  G     P  V  W     L  E  Q     L  T  K  Q     F  P  G  .
301  TGAAGGCCGC TTTAGCAATG GACCGGTCTG GCTGGAACAA CTGACGAAAC AATTTCCGGG
     ACTTCCGGCG AAATCGTTAC CTGGCCAGAC CGACCTTGTT GACTGCTTTG TTAAAGGCCC
     .  L  T  I     A  N  E  A     E  G  G     A  T  A     V  A  Y  N     K  I  S  .
361  ACTGACGATC GCTAATGAAG CAGAAGGAGG AGCAACAGCG GTCGCCTATA ACAAAATCAG
     TGACTGCTAG CGATTACTTC GTCTTCCTCC TCGTTGTCGC CAGCGGATAT TGTTTTAGTC
     .  W  D  P     K  Y  Q  V     I  N  N     L  D  Y     E  V  T  Q     F  L  Q  .
421  CTGGGACCCG AAATATCAGG TCATCAACAA CCTGGACTAT GAAGTCACAC AGTTTCTTCA
     GACCCTGGGC TTTATAGTCC AGTAGTTGTT GGACCTGATA CTTCAGTGTG TCAAAGAAGT
     .  K  D  S     F  K  P  D     D  L  V     I  L  W     V  G  A  N     D  Y  L  .
481  GAAAGACAGC TTTAAACCGG ATGATCTGGT CATCCTTTGG GTCGGCGCCA ATGATTATCT
     CTTTCTGTCG AAATTTGGCC TACTAGACCA GTAGGAAACC CAGCCGCGGT TACTAATAGA
     .  A  Y  G     W  N  T  E     Q  D  A     K  R  V     R  D  A  I     S  D  A  .
541  GGCGTATGGC TGGAACACAG AACAAGATGC CAAAAGAGTC AGAGATGCCA TCAGCGATGC
     CCGCATACCG ACCTTGTGTC TTGTTCTACG GTTTTCTCAG TCTCTACGGT AGTCGCTACG
     .  A  N  R     M  V  L  N     G  A  K     Q  I  L     L  F  N  L     P  D  L  .
601  CGCTAATAGA ATGGTCCTGA ACGGCGCCAA ACAAATCCTG CTGTTTAACC TGCCGGATCT
     GCGATTATCT TACCAGGACT TGCCGCGGTT TGTTTAGGAC GACAAATTGG ACGGCCTAGA
     .  G  Q  N     P  S  A  R     S  Q  K     V  V  E     A  V  S  H     V  S  A  .
661  GGGACAAAAT CCGAGCGCCA GAAGCCAAAA AGTCGTCGAA GCAGTCAGCC ATGTCAGCGC
     CCCTGTTTTA GGCTCGCGGT CTTCGGTTTT TCAGCAGCTT CGTCAGTCGG TACAGTCGCG
     .  Y  H  N     K  L  L  L     N  L  A     R  Q  L     A  P  T  G     M  V  K  .
721  CTATCATAAC AAACTGCTGC TGAACCTGGC AAGACAATTG GCACCGACGG GAATGGTTAA
     GATAGTATTG TTTGACGACG ACTTGGACCG TTCTGTTAAC CGTGGCTGCC CTTACCAATT
     .  L  F  E     I  D  K  Q     F  A  E     M  L  R     D  P  Q  N     F  G  L  .
781  ATTGTTTGAA ATTGACAAAC AGTTTGCCGA AATGCTGAGA GATCCGCAAA ATTTTGGCCT
     TAACAAACTT TAACTGTTTG TCAAACGGCT TTACGACTCT CTAGGCGTTT TAAAACCGGA
     .  S  D  V     E  N  P  C     Y  D  G     G  Y  V     W  K  P  F     A  T  R  .
841  GAGCGATGTC GAAAACCCGT GCTATGATGG CGGATATGTC TGGAAACCGT TTGCCACAAG
     CTCGCTACAG CTTTTGGGCA CGATACTACC GCCTATACAG ACCTTTGGCA AACGGTGTTC
     .  S  V  S     T  D  R  Q     L  S  A     F  S  P     Q  E  R  L     A  I  A  .
901  AAGCGTCAGC ACGGATAGAC AACTGTCAGC GTTTAGCCCG CAAGAAAGAC TGGCAATCGC
     TTCGCAGTCG TGCCTATCTG TTGACAGTCG CAAATCGGGC GTTCTTTCTG ACCGTTAGCG
     .  G  N  P     L  L  A  Q     A  V  A     S  P  M     A  R  R  S     A  S  P  .
961  CGGAAATCCG CTTTTGGCAC AAGCAGTTGC TTCACCGATG GCAAGAAGAT CAGCAAGCCC
     GCCTTTAGGC GAAAACCGTG TTCGTCAACG AAGTGGCTAC CGTTCTTCTA GTCGTTCGGG
     .  L  N  C     E  G  K  M     F  W  D     Q  V  H     P  T  T  V     V  H  A  .
1021 GCTGAATTGC GAAGGCAAAA TGTTTTGGGA TCAGGTCCAT CCGACAACAG TTGTCCATGC
     CGACTTAACG CTTCCGTTTT ACAAAACCCT AGTCCAGGTA GGCTGTTGTC AACAGGTACG
     .  A  L  S     E  R  A  A     T  F  I     E  T  Q     Y  E  F  L     A  H  G  .
1081 TGCCCTTTCA GAAAGAGCGG CGACGTTTAT CGAAACACAG TATGAATTTC TGGCCCATGG
     ACGGGAAAGT CTTTCTCGCC GCTGCAAATA GCTTTGTGTC ATACTTAAAG ACCGGGTACC
     .stop
1141 CTGAGTTAAC AGAGGACGGA TTTCCTGAAG GAAATCCGTT TTTTTATTTT AAGCTTGGAG
     GACTCAATTG TCTCCTGCCT AAAGGACTTC CTTTAGGCAA AAAAATAAAA TTCGAACCTC
1201 ACAAGGTAAA GGATAAAACC TCGAG
     TGTTCCATTT CCTATTTTGG AGCTC
```

FIGURE 56

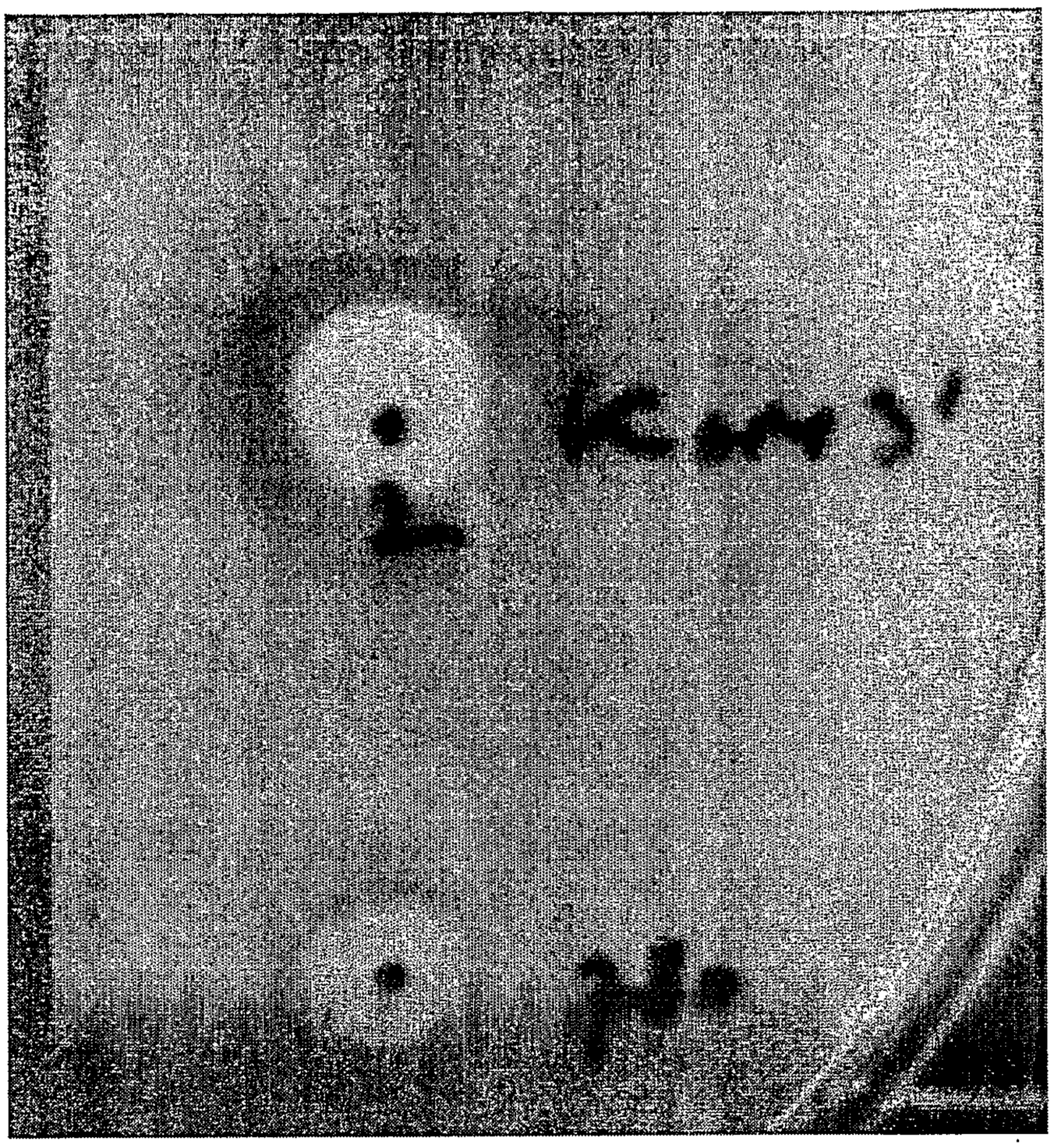


FIGURE 57 (SEQ ID No 49)

```
   1  ATGAAACAAC AAAAACGGCT TTACGCCCGA TTGCTGACGC TGTTATTTGC
      TACTTTGTTG TTTTTGCCGA AATGCGGGCT AACGACTGCG ACAATAAACG

  51  GCTCATCTTC TTGCTGCCTC ATTCTGCAGC TTCAGCAGCA GATACAAGAC
      CGAGTAGAAG AACGACGGAG TAAGACGTCG AAGTCGTCGT CTATGTTCTG

 101  CGGCGTTTAG CCGGATCGTC ATGTTTGGAG ATAGCCTGAG CGATACGGGC
      GCCGCAAATC GGCCTAGCAG TACAAACCTC TATCGGACTC GCTATGCCCG

 151  AAAATGTATA GCAAAATGAG AGGCTATCTT CCGTCAAGCC CGCCGTATTA
      TTTTACATAT CGTTTTACTC TCCGATAGAA GGCAGTTCGG GCGGCATAAT

 201  TGAAGGCCGC TTTAGCAATG GACCGGTCTG GCTGGAACAA CTGACGAAAC
      ACTTCCGGCG AAATCGTTAC CTGGCCAGAC CGACCTTGTT GACTGCTTTG

 251  AATTTCCGGG ACTGACGATC GCTAATGAAG CAGAAGGAGG AGCAACAGCG
      TTAAAGGCCC TGACTGCTAG CGATTACTTC GTCTTCCTCC TCGTTGTCGC

 301  GTCGCCTATA ACAAAATCAG CTGGGACCCG AAATATCAGG TCATCAACAA
      CAGCGGATAT TGTTTTAGTC GACCCTGGGC TTTATAGTCC AGTAGTTGTT

 351  CCTGGACTAT GAAGTCACAC AGTTTCTTCA GAAAGACAGC TTTAAACCGG
      GGACCTGATA CTTCAGTGTG TCAAAGAAGT CTTTCTGTCG AAATTTGGCC

 401  ATGATCTGGT CATCCTTTGG GTCGGCGCCA ATGATTATCT GGCGTATGGC
      TACTAGACCA GTAGGAAACC CAGCCGCGGT TACTAATAGA CCGCATACCG

 451  TGGAACACAG AACAAGATGC CAAAAGAGTC AGAGATGCCA TCAGCGATGC
      ACCTTGTGTC TTGTTCTACG GTTTTCTCAG TCTCTACGGT AGTCGCTACG

 501  CGCTAATAGA ATGGTCCTGA ACGGCGCCAA ACAAATCCTG CTGTTTAACC
      GCGATTATCT TACCAGGACT TGCCGCGGTT TGTTTAGGAC GACAAATTGG

 551  TGCCGGATCT GGGACAAAAT CCGAGCGCCA GAAGCCAAAA AGTCGTCGAA
      ACGGCCTAGA CCCTGTTTTA GGCTCGCGGT CTTCGGTTTT TCAGCAGCTT

 601  GCAGTCAGCC ATGTCAGCGC CTATCATAAC AAACTGCTGC TGAACCTGGC
      CGTCAGTCGG TACAGTCGCG GATAGTATTG TTTGACGACG ACTTGGACCG

 651  AAGACAATTG GCACCGACGG GAATGGTTAA ATTGTTTGAA ATTGACAAAC
      TTCTGTTAAC CGTGGCTGCC CTTACCAATT TAACAAACTT TAACTGTTTG

 701  AGTTTGCCGA AATGCTGAGA GATCCGCAAA ATTTTGGCCT GAGCGATGTC
      TCAAACGGCT TTACGACTCT CTAGGCGTTT TAAAACCGGA CTCGCTACAG

 751  GAAAACCCGT GCTATGATGG CGGATATGTC TGGAAACCGT TTGCCACAAG
      CTTTTGGGCA CGATACTACC GCCTATACAG ACCTTTGGCA AACGGTGTTC

 801  AAGCGTCAGC ACGGATAGAC AACTGTCAGC GTTTAGCCCG CAAGAAAGAC
      TTCGCAGTCG TGCCTATCTG TTGACAGTCG CAAATCGGGC GTTCTTTCTG

 851  TGGCAATCGC CGGAAATCCG CTTTTGGCAC AAGCAGTTGC TTCACCGATG
      ACCGTTAGCG GCCTTTAGGC GAAAACCGTG TTCGTCAACG AAGTGGCTAC

 901  GCAAGAAGAT CAGCAAGCCC GCTGAATTGC GAAGGCAAAA TGTTTTGGGA
      CGTTCTTCTA GTCGTTCGGG CGACTTAACG CTTCCGTTTT ACAAAACCCT

 951  TCAGGTCCAT CCGACAACAG TTGTCCATGC TGCCCTTTCA GAAAGAGCGG
      AGTCCAGGTA GGCTGTTGTC AACAGGTACG ACGGGAAAGT CTTTCTCGCC

1001  CGACGTTTAT CGAAACACAG TATGAATTTC TGGCCCATGG CTGA
      GCTGCAAATA GCTTTGTGTC ATACTTAAAG ACCGGGTACC GACT
```

FIGURE 58 (SEQ ID No. 50)

```
   1  ATGAAAAAAT GGTTTGTGTG TTTATTGGGA TTGGTCGCGC TGACAGTTCA GGCAGCCGAC
  61  AGCCGTCCCG CCTTCTCCCG GATCGTGATG TTTGGCGACA GCCTCTCCGA TACCGGCAAG
 121  ATGTACAGCA AGATGCGCGG TTACCTCCCC TCCAGCCCCC CCTACTATGA GGGCCGCTTC
 181  TCCAACGGGC CCGTCTGGCT GGAGCAGCTG ACCAACGAGT TCCCGGGCCT GACCATAGCC
 241  AACGAGGCGG AAGGCGGACC GACCGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
 301  TATCAGGTCA TCAACAACCT GGACTACGAG GTCACCCAGT TCCTGCAAAA AGACAGCTTC
 361  AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGCGCCAACG ACTATCTGGC CTATGGCTGG
 421  AACACAGAGC AGGATGCCAA GCGGGTGCGC GACGCCATCA GCGATGCGGC CAACCGCATG
 481  GTGCTGAACG GCGCCAAGGA GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCC
 541  TCGGCCCGCA GCCAGAAGGT GGTCGAGGCG GCCAGCCATG TCTCCGCCTA CCACAACCAG
 601  CTGCTGCTGA ACCTGGCACG CCAGCTGGCT CCCACCGGCA TGGTGAAGCT GTTCGAGATC
 661  GACAAGCAGT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT TCGGCCTGAG CGACCAGAGG
 721  AACGCCTGCT ACGGTGGCAG CTATGTATGG AAGCCGTTTG CCTCCCGCAG CGCCAGCACC
 781  GACAGCCAGC TCTCCGCCTT CAACCCGCAG GAGCGCCTCG CCATCGCCGG CAACCCGCTG
 841  CTGGCCCAGG CCGTCGCCAG CCCCATGGCT GCCCGCAGCG CCAGCACCCT CAACTGTGAG
 901  GGCAAGATGT TCTGGGATCA GGTCCACCCC ACCACTGTCG TGCACGCCGC CCTGAGCGAG
 961  CCCGCCGCCA CCTTCATCGA GAGCCAGTAC GAGTTCCTCG CCCAC
```

FIGURE 59 (SEQ ID No. 51)

```
   1  ATGAAAAAAT GGTTTGTTTG TTTATTGGGG TTGATCGCGC TGACAGTTCA GGCAGCCGAC
  61  ACTCGCCCCG CCTTCTCCCG GATCGTGATG TTCGGCGACA GCCTCTCCGA TACCGGCAAA
 121  ATGTACAGCA AGATGCGCGG TTACCTCCCC TCCAGCCCGC CCTACTATGA GGGCCGTTTC
 181  TCCAACGGAC CCGTCTGGCT GGAGCAGCTG ACCAAGCAGT TCCCGGGTCT GACCATCGCC
 241  AACGAAGCGG AAGGCGGTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
 301  TATCAGGTCT ACAACAACCT GGACTACGAG GTCACCCAGT TCTTGCAGAA AGACAGCTTC
 361  AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG ACTATCTGGC ATATGGCTGG
 421  AATACGGAGC AGGATGCCAA GCGAGTTCGC GATGCCATCA GCGATGCGGC CAACCGCATG
 481  GTACTGAACG GTGCCAAGCA GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG
 541  TCAGCCCGCA GTCAGAAGGT GGTCGAGGCG GTCAGCCATG TCTCCGCCTA TCACAACAAG
 601  CTGCTGCTGA ACCTGGCACG CCAGCTGGCC CCCACCGGCA TGGTAAAGCT GTTCGAGATC
 661  GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT TCGGCCTGAG CGACGTCGAG
 721  AACCCCTGCT ACGACGGCGG CTATGTGTGG AAGCCGTTTG CCACCCGCAG CGTCAGCACC
 781  GACCGCCAGC TCTCCGCCTT CAGTCCGCAG GAACGCCTCG CCATCGCCGG CAACCCGCTG
 841  CTGGCACAGG CCGTTGCCAG TCCTATGGCC CGCCGCAGCG CCAGCCCCCT CAACTGTGAG
 901  GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTCG TGCACGCAGC CCTGAGCGAG
 961  CGCGCCGCCA CCTTCATCGA GACCCAGTAC GAGTTCCTCG CCCACGGATG A
```

FIGURE 60 (SEQ ID No. 52)

```
   1  ATGCCGAAGC CTGCCCTTCG CCGTGTCATG ACCGCGACAG TCGCCGCCGT CGGCACGCTC
  61  GCCCTCGGCC TCACCGACGC CACCGCCCAC GCCGCGCCCG CCCAGGCCAC TCCGACCCTG
 121  GACTACGTCG CCCTCGGCGA CAGCTACAGC GCCGGCTCCG GCGTCCTGCC CGTCGACCCC
 181  GCCAACCTGC TCTGTCTGCG CTCGACGGCC AACTACCCCC ACGTCATCGC GGACACGACG
 241  GGCGCCCGCC TCACGGACGT CACCTGCGGC GCCGCGCAGA CCGCCGACTT CACGCGGGCC
 301  CAGTACCCGG GCGTCGCACC CCAGTTGGAC GCGCTCGGCA CCGGCACGGA CCTGGTCACG
 361  CTCACCATCG GCGGCAACGA CAACAGCACC TTCATCAACG CCATCACGGC CTGCGGCACG
 421  GCGGGTGTCC TCAGCGGCGG CAAGGGCAGC CCCTGCAAGG ACAGGCACGG CACCTCCTTC
 481  GACGACGAGA TCGAGGCCAA CACGTACCCC GCGCTCAAGG AGGCGCTGCT CGGCGTCCGC
 541  GCCAGGGCTC CCCACGCCAG GGTGGCGGCT CTCGGCTACC CGTGGATCAC CCCGGCCACC
 601  GCCGACCCGT CCTGCTTCCT GAAGCTCCCC CTCGCCGCCG GTGACGTGCC CTACCTGCGG
 661  GCCATCCAGG CACACCTCAA CGACGCGGTC CGGCGGGCCG CCGAGGAGAC CGGAGCCACC
 721  TACGTGGACT TCTCCGGGGT GTCCGACGGC CACGACGCCT GCGAGGCCCC CGGCACCCGC
 781  TGGATCGAAC CGCTGCTCTT CGGGCACAGC CTCGTTCCCG TCCACCCCAA CGCCCTGGGC
 841  GAGCGGCGCA TGGCCGAGCA CACGATGGAC GTCCTCGGCC TGGACTGA
```

FIGURE 61 (SEQ ID No. 53)

```
   1  TCAGTCCAGG CCGAGGACGT CCATCGTGTG CTCGGCCATG CGCCGCTCGC CCAGGGCGTT
  61  GGGGTGGACG GGAACGAGGC TGTGCCCGAA GAGCAGCGGT TCGATCCAGC GGGTGCCGGG
 121  GGCCTCGCAG GCGTCGTGGC CGTCGGACAC CCCGGAGAAG TCCACGTAGG TGGCTCCGGT
 181  CTCCTCGGCG GCCCGCCGGA CCGCGTCGTT GAGGTGTGCC TGGATGGCCC GCAGGTAGGG
 241  CACGTCACCG GCGGCGAGGG GGAGCTTCAG GAAGCAGGAC GGGTCGGCGG TGGCCGGGGT
 301  GATCCACGGG TAGCCGAGAG CCGCCACCCT GGCGTGGGGA GCCCTGGCGC GGACGCCGAG
 361  CAGCGCCTCC TTGAGCGCGG GGTACGTGTT GGCCTCGATC TCGTCGTCGA AGGAGGTGCC
 421  GTGCCTGTCC TTGCAGGGGC TGCCCTTGCC GCCGCTGAGG ACACCCGCCG TGCCGCAGGC
 481  CGTGATGGCG TTGATGAAGG TGCTGTTGTC GTTGCCGCCG ATGGTGAGCG TGACCAGGTC
 541  CGTGCCGGTG CCGAGCGCGT CCAACTGGGG TGCGACGCCC GGGTACTGGG CCCGCGTGAA
 601  GTCGGCGGTC TGCGCGGCGC CGCAGGTGAC GTCCGTGAGG CGGGCGCCCG TCGTGTCCGC
 661  GATGACGTGG GGGTAGTTGG CCGTCGAGCG CAGACAGAGC AGGTTGGCGG GGTCGACGGG
 721  CAGGACGCCG GAGCCGGCGC TGTAGCTGTC GCCGAGGGCG ACGTAGTCCA GGGTCGGAGT
 781  GGCCTGGGCG GGCGCGGCGT GGGCGGTGGC GTCGGTGAGG CCGAGGGCGA GCGTGCCGAC
 841  GGCGGCGACT GTCGCGGTCA TGACACGGCG AAGGGCAGGC TTCGGCAT
```

FIGURE 62 (SEQ ID No. 54)

```
   1  ATGGATTACG AGAAGTTTCT GTTATTTGGG GATTCCATTA CTGAATTTGC TTTTAATACT
  61  AGGCCCATTG AAGATGGCAA AGATCAGTAT GCTCTTGGAG CCGCATTAGT CAACGAATAT
 121  ACGAGAAAAA TGGATATTCT TCAAAGAGGG TTCAAAGGGT ACACTTCTAG ATGGGCGTTG
 181  AAAATACTTC CTGAGATTTT AAAGCATGAA TCCAATATTG TCATGGCCAC AATATTTTTG
 241  GGTGCCAACG ATGCATGCTC AGCAGGTCCC CAAAGTGTCC CCCTCCCCGA ATTTATCGAT
 301  AATATTCGTC AAATGGTATC TTTGATGAAG TCTTACCATA TCCGTCCTAT TATAATAGGA
 361  CCGGGGCTAG TAGATAGAGA GAAGTGGGAA AAAGAAAAAT CTGAAGAAAT AGCTCTCGGA
 421  TACTTCCGTA CCAACGAGAA CTTTGCCATT TATTCCGATG CCTTAGCAAA ACTAGCCAAT
 481  GAGGAAAAAG TTCCCTTCGT GGCTTTGAAT AAGGCGTTTC AACAGGAAGG TGGTGATGCT
 541  TGGCAACAAC TGCTAACAGA TGGACTGCAC TTTTCCGGAA AAGGGTACAA AATTTTTCAT
 601  GACGAATTAT TGAAGGTCAT TGAGACATTC TACCCCCAAT ATCATCCCAA AAACATGCAG
 661  TACAAACTGA AAGATTGGAG AGATGTGCTA GATGATGGAT CTAACATAAT GTCTTGA
```

FIGURE 63 (SEQ ID No. 55)

```
atgaacctgc gtcaatggat gggcgccgcc acggctgccc ttgccttggg cttggccgcg      60
tgcgggggcg gtgggaccga ccagagcggc aatcccaatg tcgccaaggt gcagcgcatg     120
gtggtgttcg gcgacagcct gagcgatatc ggcacctaca cccccgtcgc gcaggcggtg     180
ggcggcggca agttcaccac caacccgggc ccgatctggg ccgagaccgt ggccgcgcaa     240
ctgggcgtga cgctcacgcc ggcggtgatg ggctacgcca cctccgtgca gaattgcccc     300
aaggccggct gcttcgacta tgcgcagggc ggctcgcgcg tgaccgatcc gaacggcatc     360
ggccacaacg gcggcgcggg ggcgctgacc tacccggttc agcagcagct cgccaacttc     420
tacgcggcca gcaacaacac attcaacggc aataacgatg tcgtcttcgt gctggccggc     480
agcaacgaca ttttcttctg gaccactgcg gcggccacca gcggctccgg cgtgacgccc     540
gccattgcca cggcccaggt gcagcaggcc gcgacggacc tggtcggcta tgtcaaggac     600
atgatcgcca agggtgcgac gcaggtctac gtgttcaacc tgcccgacag cagcctgacg     660
ccggacggcg tggcaagcgg cacgaccggc caggcgctgc tgcacgcgct ggtgggcacg     720
ttcaacacga cgctgcaaag cgggctggcc ggcacctcgg cgcgcatcat cgacttcaac     780
gcacaactga ccgcggcgat ccagaatggc gcctcgttcg gcttcgccaa caccagcgcc     840
cgggcctgcg acgccaccaa gatcaatgcc ctggtgccga gcgccggcgg cagctcgctg     900
ttctgctcgg ccaacacgct ggtggcttcc ggtgcggacc agagctacct gttcgccgac     960
ggcgtgcacc cgaccacggc cggccatcgc ctgatcgcca gcaacgtgct ggcgcgcctg    1020
ctggcggata acgtcgcgca ctga                                           1044
```

FIGURE 64 (SEQ ID No. 56)

```
   1 gtgatcgggt cgtacgtggc ggtgggggac agcttcaccg agggcgtcgg cgaccccggc
  61 cccgacgggg cgttcgtcgg ctgggccgac cggctcgccg tactgctcgc ggaccggcgc
 121 cccgagggcg acttcacgta cacgaacctc gccgtgcgcg gcaggctcct cgaccagatc
 181 gtggcggaac aggtcccgcg ggtcgtcgga ctcgcgcccg acctcgtctc gttcgcggcg
 241 ggcggcaacg acatcatccg gcccggcacc gatcccgacg aggtcgccga gcggttcgag
 301 ctggcggtgg ccgcgctgac cgccgcggcc ggaaccgtcc tggtgaccac cgggttcgac
 361 acccgggggg tgcccgtcct caagcacctg cgcggcaaga tcgccacgta caacgggcac
 421 gtccgcgcca tcgccgaccg ctacggctgc ccggtgctcg acctgtggtc gctgcggagc
 481 gtccaggacc gcagggcgtg ggacgccgac cggctgcacc tgtcgccgga ggggcacacc
 541 cgggtggcgc tgcgcgcggg gcaggccctg ggcctgcgcg tcccggccga ccctgaccag
 601 ccctggccgc ccctgccgcc gcgcggcacg ctcgacgtcc ggcgcgacga cgtgcactgg
 661 gcgcgcgagt acctggtgcc gtggatcggg cgccggctgc ggggcgagtc gtcgggcgac
 721 cacgtgacgg ccaaggggac gctgtcgccg gacgccatca agacgcggat cgccgcggtg
 781 gcctga
```

FIGURE 65 (SEQ ID No. 57)

```
   1 atgcagacga accccgcgta caccagtctc gtcgccgtcg gcgactcctt caccgagggc
  61 atgtcggacc tgctgcccga cggctcctac cgtggctggg ccgacctcct cgccacccgg
 121 atggcggccc gctcccccgg cttccggtac gccaacctgg cggtgcgcgg gaagctgatc
 181 ggacagatcg tcgacgagca ggtggacgtg gccgccgcca tgggagccga cgtgatcacg
 241 ctggtcggcg ggctcaacga cacgctgcgg cccaagtgcg acatggcccg ggtgcgggac
 301 ctgctgaccc aggccgtgga acggctcgcc ccgcactgcg agcagctggt gctgatgcgc
 361 agtcccggtc gccagggtcc ggtgctggag cgcttccggc cccgcatgga ggccctgttc
 421 gccgtgatcg acgacctggc cgggcggcac ggcgccgtgg tcgtcgacct gtacggggcc
 481 cagtcgctgg ccgaccctcg gatgtgggac gtggaccggc tgcacctgac cgccgagggc
 541 caccgccggg tcgcggaggc ggtgtggcag tcgctcggcc acgagcccga ggaccccgag
 601 tggcacgcgc cgatcccggc gacgccgccg ccggggtggg tgacgcgcag gaccgcggac
 661 gtccggttcg cccggcagca cctgctgccc tggataggcc gcaggctgac cgggcgctcg
 721 tccggggacg gcctgccggc caagcgcccg gacctgctgc cctacgagga ccccgcacgg
 781 tga
```

FIGURE 66 (SEQ ID No. 58)

```
   1 atgacccggg gtcgtgacgg gggtgcgggg gcgcccccca ccaagcaccg tgccctgctc
  61 gcggcgatcg tcaccctgat agtggcgatc tccgcggcca tatacgccgg agcgtccgcg
 121 gacgacggca gcagggacca cgcgctgcag gccggaggcc gtctcccacg aggagacgcc
 181 gccccccgcgt ccaccggtgc ctgggtgggc gcctgggcca ccgcaccggc cgcggccgag
 241 ccgggcaccg agacgaccgg cctggcgggc cgctccgtgc gcaacgtcgt gcacacctcg
 301 gtcggcggca ccggcgcgcg gatcaccctc tcgaacctgt acgggcagtc gccgctgacc
 361 gtcacacacg cctcgatcgc cctggccgcc gggcccgaca ccgccgccgc gatcgccgac
 421 accatgcgcc ggctcacctt cggcggcagc gcccgggtga tcatcccggc gggcggccag
 481 gtgatgagcg acaccgcccg cctcgccatc ccctacgggg cgaacgtcct ggtcaccacg
 541 tactccccca tcccgtccgg gccggtgacc taccatccgc aggcccggca gaccagctac
 601 ctggccgacg gcgaccgcac ggcggacgtc accgccgtcg cgtacaccac ccccacgccc
 661 tactggcgct acctgaccgc cctcgacgtg ctgagccacg aggccgacgg cacggtcgtg
 721 gcgttcggcg actccatcac cgacggcgcc cgctcgcaga gcgacgccaa ccaccgctgg
 781 accgacgtcc tcgccgcacg cctgcacgag gcggcgggcg acggccggga cacgccccgc
 841 tacagcgtcg tcaacgaggg catcagcggc aaccggctcc tgaccagcag gccggggcgg
 901 ccggccgaca acccgagcgg actgagccgg ttccagcggg acgtgctgga acgcaccaac
 961 gtcaaggccg tcgtcgtcgt cctcggcgtc aacgacgtcc tgaacagccc ggaactcgcc
1021 gaccgcgacg ccatcctgac cggcctgcgc accctcgtcg accgggcgca cgcccgggga
1081 ctgcgggtcg tcggcgccac gatcacgccg ttcggcggct acggcggcta caccgaggcc
1141 cgcgagacga tgcggcagga ggtcaacgag gagatccgct ccggccgggt cttcgacacg
1201 gtcgtcgact tcgacaaggc cctgcgcgac ccgtacgacc cgcgccggat gcgctccgac
1261 tacgacagcg gcgaccacct gcaccccggc gacaaggggt acgcgcgcat gggcgcggtc
1321 atcgacctgg ccgcgctgaa gggcgcggcg ccggtcaagg cgtag
```

FIGURE 67 (SEQ ID No. 59)

```
   1 atgacgagca tgtcgagggc gagggtggcg cggcggatcg cggccggcgc ggcgtacggc
  61 ggcggcggca tcggcctggc gggagcggcg gcggtcggtc tggtggtggc cgaggtgcag
 121 ctggccagac gcagggtggg ggtgggcacg ccgacccggg tgccgaacgc gcagggactg
 181 tacggcggca ccctgcccac ggccggcgac ccgccgctgc ggctgatgat gctgggcgac
 241 tccacggccg ccgggcaggg cgtgcaccgg gccgggcaga cgccgggcgc gctgctggcg
 301 tccgggctcg cggcggtggc ggagcggccg gtgcggctgg ggtcggtcgc ccagccgggg
 361 gcgtgctcgg acgacctgga ccggcaggtg gcgctggtgc tcgccgagcc ggaccgggtg
 421 cccgacatct gcgtgatcat ggtcggcgcc aacgacgtca cccaccggat gccggcgacc
 481 cgctcggtgc ggcacctgtc ctcggcggta cggcggctgc gcacggccgg tgcggaggtg
 541 gtggtcggca cctgtccgga cctgggcacg atcgagcggg tgcggcagcc gctgcgctgg
 601 ctggcccggc gggcctcacg gcagctcgcg gcggcacaga ccatcggcgc cgtcgagcag
 661 ggcgggcgca cggtgtcgct gggcgacctg ctgggtccgg agttcgcgca gaacccgcgg
 721 gagctcttcg gccccgacaa ctaccacccc tccgccgagg ggtacgccac ggccgcgatg
 781 gcggtactgc cctcggtgtg cgccgcgctc ggcctgtggc cggccgacga ggagcacccg
 841 gacgcgctgc gccgcgaggg cttcctgccg gtggcgcgcg cggcggcgga ggcggcgtcc
 901 gaggcgggta cggaggtcgc cgccgccatg cctacggggc ctcgggggcc ctgggcgctg
 961 ctgaagcgcc ggagacggcg tcgggtgtcg gaggcggaac cgtccagccc gtccggcgtt
1021 tga
```

FIGURE 68 (SEQ ID No. 60)

```
   1 atgggtcgag ggacggacca gcggacgcgg tacggccgtc gccgggcgcg tgtcgcgctc
  61 gccgccctga ccgccgccgt cctgggcgtg ggcgtggcgg gctgcgactc cgtgggcggc
 121 gactcacccg ctccttccgg cagcccgtcg aagcggacga ggacggcgcc cgcctgggac
 181 accagcccgg cgtccgtcgc cgccgtgggc gactccatca cgcgcggctt cgacgcctgt
 241 gcggtgctgt cggactgccc ggaggtgtcg tgggcgaccg gcagcagcgc gaaggtcgac
 301 tcgctggccg tacggctgct ggggaaggcg gacgcggccg agcacagctg gaactacgcg
 361 gtcaccgggg cccggatggc ggacctgacc gctcaggtga cgcgggcggc gcagcgcgag
 421 ccggagctgg tggcggtgat ggccggggcg aacgacgcgt gccggtccac gacctcggcg
 481 atgacgccgg tggcggactt ccgggcgcag ttcgaggagg cgatggccac cctgcgcaag
 541 aagctcccca aggcgcaggt gtacgtgtcg agcatcccgg acctcaagcg gctctggtcc
 601 cagggccgca ccaacccgct gggcaagcag gtgtggaagc tcggcctgtg cccgtcgatg
 661 ctgggcgacg cggactccct ggactcggcg gcgaccctgc ggcgcaacac ggtgcgcgac
 721 cgggtggcgg actacaacga ggtgctgcgg gaggtctgcg cgaaggaccg gcggtgccgc
 781 agcgacgacg gcgcggtgca cgagttccgg ttcggcacgg accagttgag ccactgggac
 841 tggttccacc cgagtgtgga cggccaggcc cggctggcgg agatcgccta ccgcgcggtc
 901 accgcgaaga atccctga
```

FIGURE 69 (SEQ ID No. 61)

```
   1 ttcatcacaa cgatgtcaca acaccggcca tccgggtcat ccctgatcgt gggaatgggt
  61 gacaagcctt cccgtgacga aagggtcctg ctacatcaga aatgacagaa atcctgctca
 121 gggaggttcc atgagactgt cccgacgcgc ggccacggcg tccgcgctcc tcctcacccc
 181 ggcgctcgcg ctcttcggcg cgagcgccgc cgtgtccgcg ccgcgaatcc aggccaccga
 241 ctacgtggcc ctcggcgact cctactcctc gggggtcggc gcgggcagct acgacagcag
 301 cagtggctcc tgtaagcgca gcaccaagtc ctacccggcc ctgtgggccg cctcgcacac
 361 cggtacgcgg ttcaacttca ccgcctgttc gggcgcccgc acaggagacg tgctggccaa
 421 gcagctgacc ccggtcaact ccggcaccga cctggtcagc attaccatcg gcggcaacga
 481 cgcgggcttc gccgacacca tgaccacctg caacctccag ggcgagagcg cgtgcctggc
 541 gcggatcgcc aaggcgcgcg cctacatcca gcagacgctg cccgcccagc tggaccaggt
 601 ctacgacgcc atcgacagcc gggcccccgc agcccaggtc gtcgtcctgg gctacccgcg
 661 cttctacaag ctgggcggca gctgcgccgt cggtctctcg gagaagtccc gcgcggccat
 721 caacgccgcc gccgacgaca tcaacgccgt caccgccaag cgcgccgccg accacggctt
 781 cgccttcggg gacgtcaaca cgaccttcgc cgggcacgag ctgtgctccg gcgccccctg
 841 gctgcacagc gtcaccttc ccgtggagaa ctcctaccac cccacggcca acggacagtc
 901 caagggctac ctgcccgtcc tgaactccgc cacctgatct cgcggctact ccgccctga
 961 cgaagtcccg ccccgggcg gggcttcgcc gtaggtgcgc gtaccgccgt cgcccgtcgc
1021 gccggtggcc ccgccgtacg tgccgccgcc cccggacgcg gtcggttc
```

FIGURE 70 (SEQ ID No. 62)

```
   1  ATGAAAAAAT GGTTTGTGTG TTTATTGGGA TTGGTCGCGC TGACAGTTCA
      TACTTTTTTA CCAAACACAC AAATAACCCT AACCAGCGCG ACTGTCAAGT

  51  GGCAGCCGAC AGTCGCCCCG CCTTTTCCCG GATCGTGATG TTCGGCGACA
      CCGTCGGCTG TCAGCGGGGC GGAAAAGGGC CTAGCACTAC AAGCCGCTGT

 101  GCCTCTCCGA TACCGGCAAA ATGTACAGCA AGATGCGCGG TTACCTCCCC
      CGGAGAGGCT ATGGCCGTTT TACATGTCGT TCTACGCGCC AATGGAGGGG

 151  TCCAGCCCGC CCTACTATGA GGGCCGTTTC TCCAACGGAC CCGTCTGGCT
      AGGTCGGGCG GGATGATACT CCCGGCAAAG AGGTTGCCTG GGCAGACCGA

 201  GGAGCAGCTG ACCAAACAGT TCCCGGGTCT GACCATCGCC AACGAAGCGG
      CCTCGTCGAC TGGTTTGTCA AGGGCCCAGA CTGGTAGCGG TTGCTTCGCC

 251  AAGGCGGTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
      TTCCGCCACG GTGACGGCAC CGAATGTTGT TCTAGAGGAC CTTAGGGTTC

 301  TATCAGGTCA TCAACAACCT GGACTACGAG GTCACCCAGT TCTTGCAGAA
      ATAGTCCAGT AGTTGTTGGA CCTGATGCTC CAGTGGGTCA AGAACGTCTT

 351  AGACAGCTTC AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG
      TCTGTCGAAG TTCGGCCTGC TAGACCACTA GGAGACCCAG CCACGGTTAC

 401  ACTATCTGGC CTATGGCTGG AACACGGAGC AGGATGCCAA GCGGGTTCGC
      TGATAGACCG GATACCGACC TTGTGCCTCG TCCTACGGTT CGCCCAAGCG

 451  GATGCCATCA GCGATGCGGC CAACCGCATG GTACTGAACG GTGCCAAGCA
      CTACGGTAGT CGCTACGCCG GTTGGCGTAC CATGACTTGC CACGGTTCGT

 501  GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG TCAGCTCGCA
      CTATGACGAC AAGTTGGACG GCCTAGACCC GGTCTTGGGC AGTCGAGCGT

 551  GTCAGAAGGT GGTCGAGGCG GTCAGCCATG TCTCCGCCTA TCACAACCAG
      CAGTCTTCCA CCAGCTCCGC CAGTCGGTAC AGAGGCGGAT AGTGTTGGTC

 601  CTGCTGCTGA ACCTGGCACG CCAGCTGGCC CCCACCGGCA TGGTAAAGCT
      GACGACGACT TGGACCGTGC GGTCGACCGG GGGTGGCCGT ACCATTTCGA

 651  GTTCGAGATC GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT
      CAAGCTCTAG CTGTTCGTTA AACGGCTCTA CGACGCACTA GGCGTCTTGA

 701  TCGGCCTGAG CGACGTCGAG AACCCCTGCT ACGACGGCGG CTATGTGTGG
      AGCCGGACTC GCTGCAGCTC TTGGGGACGA TGCTGCCGCC GATACACACC

 751  AAGCCGTTTG CCACCCGCAG CGTCAGCACC GACCGCCAGC TCTCCGCCTT
      TTCGGCAAAC GGTGGGCGTC GCAGTCGTGG CTGGCGGTCG AGAGGCGGAA

 801  CAGTCCGCAG GAACGCCTCG CCATCGCCGG CAACCCGCTG CTGGCACAGG
      GTCAGGCGTC CTTGCGGAGC GGTAGCGGCC GTTGGGCGAC GACCGTGTCC

 851  CCGTTGCCAG TCCTATGGCC CGCCGCAGCG CCAGCCCCCT CAACTGTGAG
      GGCAACGGTC AGGATACCGG GCGGCGTCGC GGTCGGGGGA GTTGACACTC

 901  GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTCG TGCACGCAGC
      CCGTTCTACA AGACCCTAGT CCATGTGGGC TGGTGACAGC ACGTGCGTCG

 951  CCTGAGCGAG CGCGCCGCCA CCTTCATCGC GAACCAGTAC GAGTTCCTCG
      GGACTCGCTC GCGCGGCGGT GGAAGTAGCG CTTGGTCATG CTCAAGGAGC

1001  CCCAC TGA
      GGGTG ACT
```

FIGURE 71 (SEQ ID No. 63)

```
   1  ATGAAAAAAT GGTTTGTTTG TTTATTGGGG TTGATCGCGC TGACAGTTCA
      TACTTTTTTA CCAAACAAAC AAATAACCCC AACTAGCGCG ACTGTCAAGT

  51  GGCAGCCGAC ACTCGCCCCG CCTTCTCCCG GATCGTGATG TTCGGCGACA
      CCGTCGGCTG TGAGCGGGGC GGAAGAGGGC CTAGCACTAC AAGCCGCTGT

 101  GCCTCTCCGA TACCGGCAAA ATGTACAGCA AGATGCGCGG TTACCTCCCC
      CGGAGAGGCT ATGGCCGTTT TACATGTCGT TCTACGCGCC AATGGAGGGG

 151  TCCAGCCCGC CCTACTATGA GGGCCGTTTC TCCAACGGAC CCGTCTGGCT
      AGGTCGGGCG GGATGATACT CCCGGCAAAG AGGTTGCCTG GGCAGACCGA

 201  GGAGCAGCTG ACCAAGCAGT TCCCGGGTCT GACCATCGCC AACGAAGCGG
      CCTCGTCGAC TGGTTCGTCA AGGGCCCAGA CTGGTAGCGG TTGCTTCGCC

 251  AAGGCGGTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
      TTCCGCCACG GTGACGGCAC CGAATGTTGT TCTAGAGGAC CTTAGGGTTC

 301  TATCAGGTCA TCAACAACCT GGACTACGAG GTCACCCAGT TCTTGCAGAA
      ATAGTCCAGT AGTTGTTGGA CCTGATGCTC CAGTGGGTCA AGAACGTCTT

 351  AGACAGCTTC AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG
      TCTGTCGAAG TTCGGCCTGC TAGACCACTA GGAGACCCAG CCACGGTTAC

 401  ACTATCTGGC ATATGGCTGG AATACGGAGC AGGATGCCAA GCGAGTTCGC
      TGATAGACCG TATACCGACC TTATGCCTCG TCCTACGGTT CGCTCAAGCG

 451  GATGCCATCA GCGATGCGGC CAACCGCATG GTACTGAACG GTGCCAAGCA
      CTACGGTAGT CGCTACGCCG GTTGGCGTAC CATGACTTGC CACGGTTCGT

 501  GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG TCAGCCCGCA
      CTATGACGAC AAGTTGGACG GCCTAGACCC GGTCTTGGGC AGTCGGGCGT

 551  GTCAGAAGGT GGTCGAGGCG GTCAGCCATG TCTCCGCCTA TCACAACAAG
      CAGTCTTCCA CCAGCTCCGC CAGTCGGTAC AGAGGCGGAT AGTGTTGTTC

 601  CTGCTGCTGA ACCTGGCACG CCAGCTGGCC CCCACCGGCA TGGTAAAGCT
      GACGACGACT TGGACCGTGC GGTCGACCGG GGGTGGCCGT ACCATTTCGA

 651  GTTCGAGATC GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT
      CAAGCTCTAG CTGTTCGTTA AACGGCTCTA CGACGCACTA GGCGTCTTGA

 701  TCGGCCTGAG CGACGTCGAG AACCCCTGCT ACGACGGCGG CTATGTGTGG
      AGCCGGACTC GCTGCAGCTC TTGGGGACGA TGCTGCCGCC GATACACACC

 751  AAGCCGTTTG CCACCCGCAG CGTCAGCACC GACCGCCAGC TCTCCGCCTT
      TTCGGCAAAC GGTGGGCGTC GCAGTCGTGG CTGGCGGTCG AGAGGCGGAA

 801  CAGTCCGCAG GAACGCCTCG CCATCGCCGG CAACCCGCTG CTGGCACAGG
      GTCAGGCGTC CTTGCGGAGC GGTAGCGGCC GTTGGGCGAC GACCGTGTCC

 851  CCGTTGCCAG TCCTATGGCC CGCCGCAGCG CCAGCCCCCT CAACTGTGAG
      GGCAACGGTC AGGATACCGG GCGGCGTCGC GGTCGGGGGA GTTGACACTC

 901  GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTCG TGCACGCAGC
      CCGTTCTACA AGACCCTAGT CCATGTGGGC TGGTGACAGC ACGTGCGTCG

 951  CCTGAGCGAG CGCGCCGCCA CCTTCATCGA GACCCAGTAC GAGTTCCTCG
      GGACTCGCTC GCGCGGCGGT GGAAGTAGCT CTGGGTCATG CTCAAGGAGC

1001  CCCACGGATG A
      GGGTGCCTAC T
```

FIGURE 72 (SEQ ID No. 24)

```
   1  ATGTTTAAGT TTAAAAAGAA TTTCTTAGTT GGATTATCGG CAGCTTTAAT
      TACAAATTCA AATTTTTCTT AAAGAATCAA CCTAATAGCC GTCGAAATTA

  51  GAGTATTAGC TTGTTTTCGG CAACCGCCTC TGCAGCTAGC GCCGACAGCC
      CTCATAATCG AACAAAAGCC GTTGGCGGAG ACGTCGATCG CGGCTGTCGG

 101  GTCCCGCCTT TTCCCGGATC GTGATGTTCG GCGACAGCCT CTCCGATACC
      CAGGGCGGAA AAGGGCCTAG CACTACAAGC CGCTGTCGGA GAGGCTATGG

 151  GGCAAAATGT ACAGCAAGAT GCGCGGTTAC CTCCCCTCCA GCCCGCCCTA
      CCGTTTTACA TGTCGTTCTA CGCGCCAATG GAGGGGAGGT CGGGCGGGAT

 201  CTATGAGGGC CGTTTCTCCA ACGGACCCGT CTGGCTGGAG CAGCTGACCA
      GATACTCCCG GCAAAGAGGT TGCCTGGGCA GACCGACCTC GTCGACTGGT

 251  AACAGTTCCC GGGTCTGACC ATCGCCAACG AAGCGGAAGG CGGTGCCACT
      TTGTCAAGGG CCCAGACTGG TAGCGGTTGC TTCGCCTTCC GCCACGGTGA

 301  GCCGTGGCTT ACAACAAGAT CTCCTGGAAT CCCAAGTATC AGGTCATCAA
      CGGCACCGAA TGTTGTTCTA GAGGACCTTA GGGTTCATAG TCCAGTAGTT

 351  CAACCTGGAC TACGAGGTCA CCCAGTTCTT GCAGAAAGAC AGCTTCAAGC
      GTTGGACCTG ATGCTCCAGT GGGTCAAGAA CGTCTTTCTG TCGAAGTTCG

 401  CGGACGATCT GGTGATCCTC TGGGTCGGTG CCAATGACTA TCTGGCCTAT
      GCCTGCTAGA CCACTAGGAG ACCCAGCCAC GGTTACTGAT AGACCGGATA

 451  GGCTGGAACA CGGAGCAGGA TGCCAAGCGG GTTCGCGATG CCATCAGCGA
      CCGACCTTGT GCCTCGTCCT ACGGTTCGCC CAAGCGCTAC GGTAGTCGCT

 501  TGCGGCCAAC CGCATGGTAC TGAACGGTGC CAAGCAGATA CTGCTGTTCA
      ACGCCGGTTG GCGTACCATG ACTTGCCACG GTTCGTCTAT GACGACAAGT

 551  ACCTGCCGGA TCTGGGCCAG AACCCGTCAG CTCGCAGTCA GAAGGTGGTC
      TGGACGGCCT AGACCCGGTC TTGGGCAGTC GAGCGTCAGT CTTCCACCAG

 601  GAGGCGGTCA GCCATGTCTC CGCCTATCAC AACCAGCTGC TGCTGAACCT
      CTCCGCCAGT CGGTACAGAG GCGGATAGTG TTGGTCGACG ACGACTTGGA

 651  GGCACGCCAG CTGGCCCCCA CCGGCATGGT AAAGCTGTTC GAGATCGACA
      CCGTGCGGTC GACCGGGGGT GGCCGTACCA TTTCGACAAG CTCTAGCTGT

 701  AGCAATTTGC CGAGATGCTG CGTGATCCGC AGAACTTCGG CCTGAGCGAC
      TCGTTAAACG GCTCTACGAC GCACTAGGCG TCTTGAAGCC GGACTCGCTG

 751  GTCGAGAACC CCTGCTACGA CGGCGGCTAT GTGTGGAAGC CGTTTGCCAC
      CAGCTCTTGG GGACGATGCT GCCGCCGATA CACACCTTCG GCAAACGGTG

 801  CCGCAGCGTC AGCACCGACC GCCAGCTCTC CGCCTTCAGT CCGCAGGAAC
      GGCGTCGCAG TCGTGGCTGG CGGTCGAGAG GCGGAAGTCA GGCGTCCTTG

 851  GCCTCGCCAT CGCCGGCAAC CCGCTGCTGG CACAGGCCGT TGCCAGTCCT
      CGGAGCGGTA GCGGCCGTTG GGCGACGACC GTGTCCGGCA ACGGTCAGGA

 901  ATGGCCCGCC GCAGCGCCAG CCCCCTCAAC TGTGAGGGCA AGATGTTCTG
      TACCGGGCGG CGTCGCGGTC GGGGGAGTTG ACACTCCCGT TCTACAAGAC

 951  GGATCAGGTA CACCCGACCA CTGTCGTGCA CGCAGCCCTG AGCGAGCGCG
      CCTAGTCCAT GTGGGCTGGT GACAGCACGT GCGTCGGGAC TCGCTCGCGC

1001  CCGCCACCTT CATCGCGAAC CAGTACGAGT TCCTCGCCCA CTGATGA
      GGCGGTGGAA GTAGCGCTTG GTCATGCTCA AGGAGCGGGT GACTACT
```