



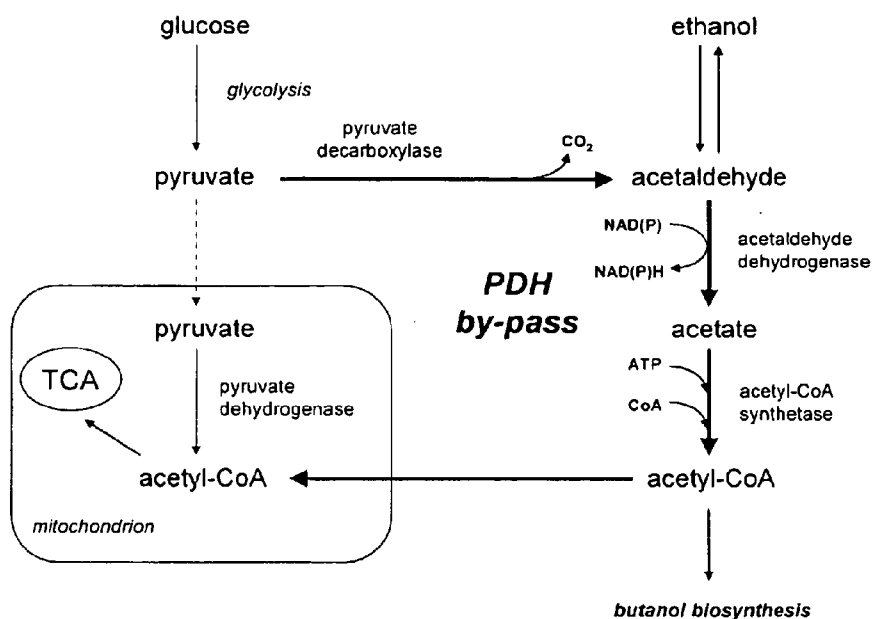
US 20100248233A1

(19) **United States**(12) **Patent Application Publication**
Müller et al.(10) **Pub. No.: US 2010/0248233 A1**(43) **Pub. Date: Sep. 30, 2010**(54) **ACETYL-CoA PRODUCING ENZYMES IN YEAST**(75) Inventors: **Ulrike Maria Müller**, Linnich (DE); **Liang Wu**, Delft (NL); **Lourina Madeleine Raamsdonk**, Nootdorp (NL); **Aaron Adriaan Winkler**, Den Haag (NL)**Publication Classification**(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C12N 15/63 (2006.01)
C12N 1/19 (2006.01)
C12P 7/16 (2006.01)Correspondence Address:
NIXON & VANDERHYE, PC
901 NORTH GLEBE ROAD, 11TH FLOOR
ARLINGTON, VA 22203 (US)(52) **U.S. Cl. .. 435/6; 435/320.1; 435/254.2; 435/254.21; 435/160**(73) Assignee: **DSM IP ASSETS B.V.**, Heerlen (NL)(21) Appl. No.: **12/670,050**(22) PCT Filed: **Jul. 11, 2008**(86) PCT No.: **PCT/EP08/59119**§ 371 (c)(1),
(2), (4) Date: **May 7, 2010****Related U.S. Application Data**

(60) Provisional application No. 60/935,031, filed on Jul. 23, 2007, provisional application No. 61/064,120, filed on Feb. 19, 2008.

(30) **Foreign Application Priority Data**Jul. 23, 2007 (EP) 07112956.3
Dec. 21, 2007 (EP) 07123976.8
Feb. 19, 2008 (EP) 08101747.7(57) **ABSTRACT**

The present invention relates to a method of identifying a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) a yeast cell comprising: a) providing a mutated yeast cell comprising a deletion of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS); b) transforming said mutated yeast cell with an expression vector comprising a heterologous nucleotide sequence encoding a candidate polypeptide having potential enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA; c) testing said recombinant mutated yeast cell for its ability to grow on minimal medium containing glucose as sole carbon source, and d) identifying said candidate polypeptide as a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell when growth of said cell is observed. The invention further relates to a method of producing a fermentation production such as butanol.



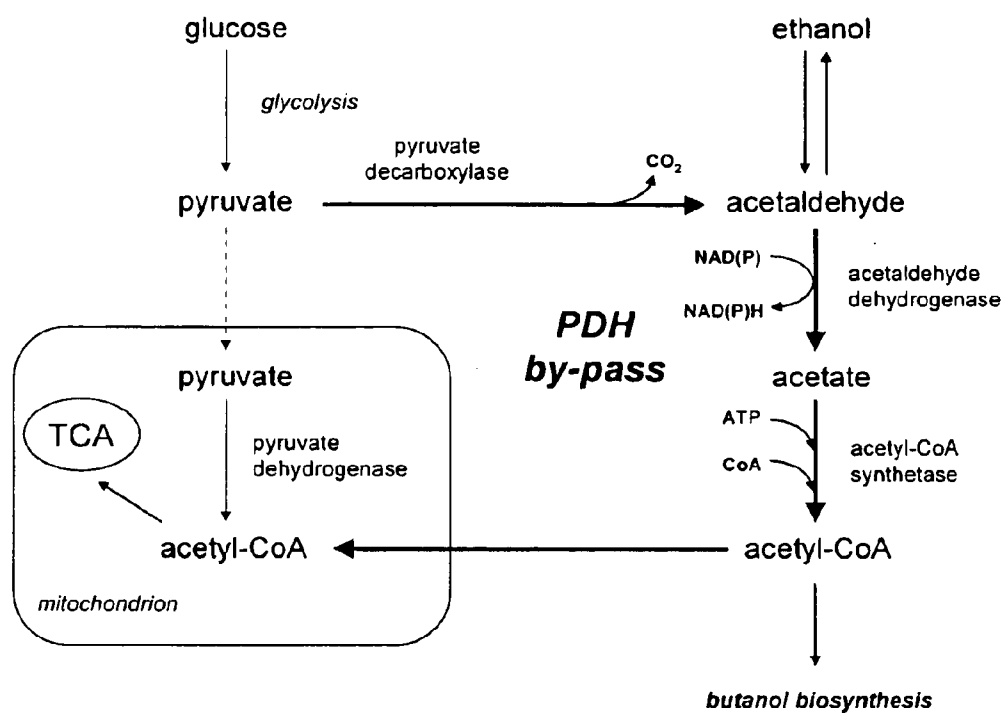


Figure 1

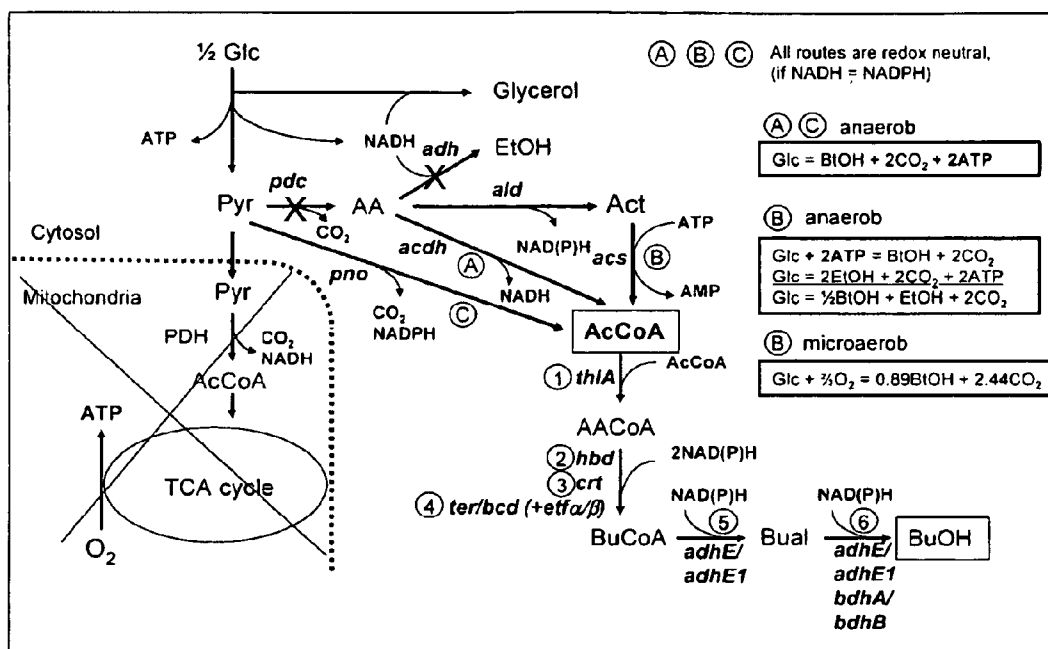


Figure 2

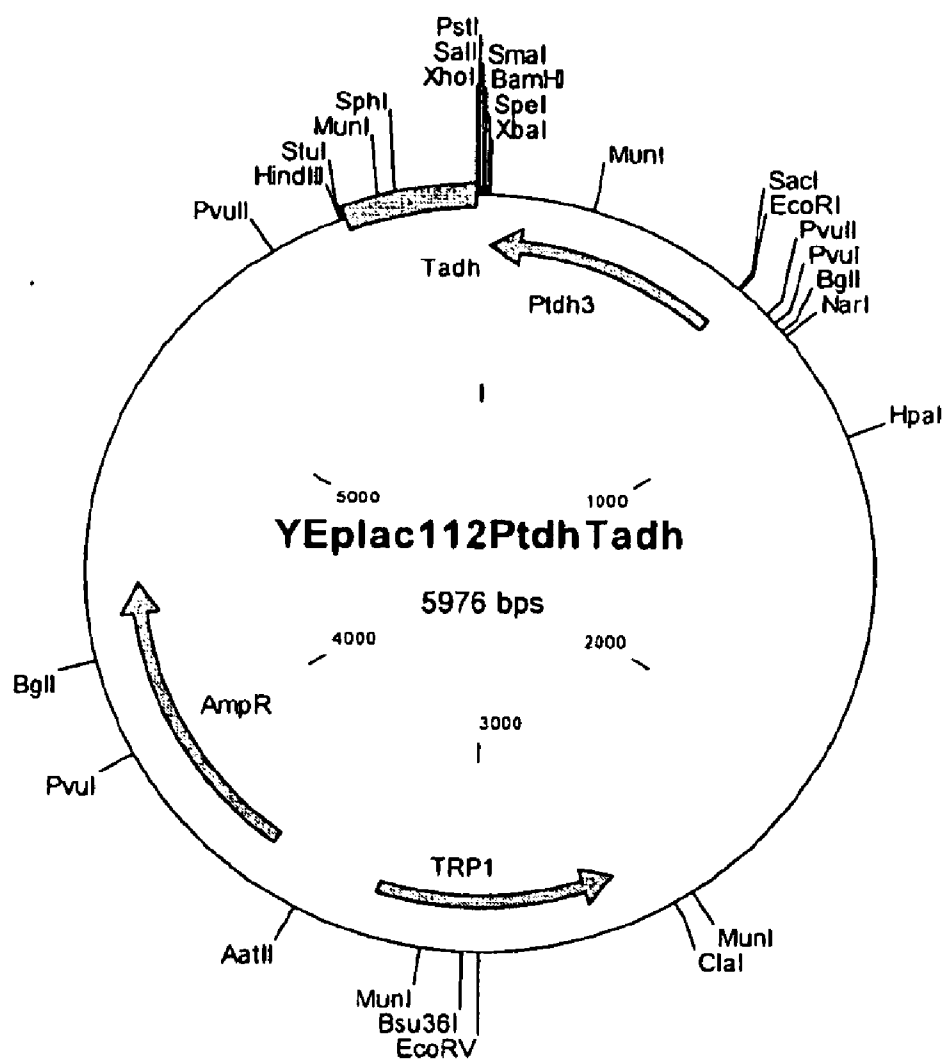


Figure 3



Figure 4

ACETYL-COA PRODUCING ENZYMES IN YEAST

FIELD OF THE INVENTION

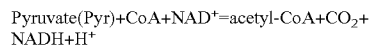
[0001] The present invention is in the field of metabolites production in yeast using heterologous expression systems. In particular, the present invention relates to the metabolic engineering of yeast strains capable of producing metabolites that require cytosolic acetyl-CoA as a precursor, such as butanol-producing yeast strains. The present invention relates to an assay system for identifying heterologous enzymes capable of converting pyruvate, acetaldehyde or acetate into cytosolic acetyl-CoA when expressed in the cytosol in yeast.

BACKGROUND OF THE INVENTION

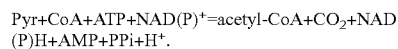
[0002] Acetyl-coenzyme A (CoA) is an essential intermediate in numerous metabolic pathways, and is a key precursor in the synthesis of many industrial relevant compounds, such as fatty acids, carotenoids, isoprenoids, vitamins, amino acids, lipids, wax esters, (poly)saccharides polyhydroxyalkanoates, statins, polyketides and acetic esters (such as ethyl acetate and isoamyl acetate). In particular, acetyl-CoA is also the precursor of the industrially important bulk chemical 1-butanol.

[0003] Compared to bacteria, such as *E. coli*, yeast cells provide a very suitable alternative to produce the above-mentioned acetyl-CoA derived products, in that yeast is not susceptible to phage or other infection since yeast-based processes may be run at low pH. Therefore, the use of yeast does not require a sterile process, thereby lowering the cost price of the product of interest.

[0004] When natural (wild type) yeast is not able to produce the acetyl-CoA-derived product of interest, the use of metabolic engineering can provide for yeast cells expressing heterologous genes that could support such a process. In such cases, the heterologous gene products are usually targeted to the cytosolic compartment of yeast. As the biosynthesis of acetyl-CoA-derived product will take place completely or partially in the cytosol, the supply of sufficient amounts of the precursor acetyl-CoA in the cytosolic compartment is crucial. In *Saccharomyces cerevisiae*, biosynthesis of acetyl-CoA takes place in two separate compartments. In mitochondria, acetyl-CoA is synthesized by oxidative decarboxylation of pyruvate catalyzed by the pyruvate dehydrogenase complex (PDH), with the following overall reaction stoichiometry:



[0005] In cytosol, acetyl-CoA is synthesized via the pyruvate dehydrogenase (PDH) by-pass, involving the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS), with the following overall reaction stoichiometry:



[0006] Pyruvate-decarboxylase-negative (Pdc-) mutant of the yeast *S. cerevisiae* does not have a functional PDH by-pass, and cannot grow on minimal medium with glucose as the sole carbon source due to inability to supply (sufficient) cytosolic acetyl-CoA for growth (Flikweert et al., (1996) Yeast 12:247-57). The PDH by-pass is therefore essential in providing acetyl-CoA in the cytosolic compartment. However, the PDH by-pass in yeast is not optimal with respect to

the energy balance, as can be seen from the overall reaction stoichiometry: 2 moles of ATP are needed per acetyl-CoA synthesized via the PDH-bypass since in the acetyl-CoA synthetase reaction ATP is hydrolyzed to AMP. In contrast, the mitochondrial pathway via the PDH requires no ATP. The additional ATP requirement of the PDH by-pass can present a problem for synthesizing the product of interest from cytosolic acetyl-CoA precursor, as more carbon source needs to be diverted for ATP generation, via e.g. oxidative phosphorylation and/or substrate phosphorylation (e.g. glycolysis), thereby lowering the overall yield of the product on carbon.

[0007] When yeast is metabolically engineered to produce 1-butanol, heterologous biosynthetic genes of 1-butanol can be expressed in the cytosol in yeast cells (WO 2007/041269). In general 1 mole of glucose give rise to 2 moles of acetyl-CoA via glycolysis, which is the precursor of 1 mole of butanol; hence a maximum of 1 mole of butanol can be synthesized per mole of glucose if cell growth and maintenance is not considered. However, when the PDH by-pass is used in combination with butanol biosynthesis, this maximal theoretical yield cannot be achieved due to energy imbalance: whereas 2 moles of ATP are generated per mole of glucose converted in glycolysis, a total of 4 moles (2 times 2 mole) of ATP are needed in the PDH by-pass to form 2 moles of acetyl-CoA, which are converted to 1 mole of butanol. Thus, there is a net shortage of ATP if the PDH by-pass were used to synthesize 1 mole of 1-butanol from 1 mole of glucose.

[0008] Thus, there is a need for the identification of possible alternative metabolic routes for producing cytosolic acetyl-CoA in yeast, for the production of acetyl-CoA-derived products, in particular butanol, wherein the PDH by-pass is not required.

[0009] Butanol is an important industrial chemical and is suitable as an alternative engine fuel having improved properties over ethanol. Butanol also finds use as a solvent for a wide variety of chemical and textile processes, in the organic synthesis of plastics, as a chemical intermediate and as a solvent in the coating and food and flavor industry. Butanol can be produced from biomass (biobutanol) as well as fossil fuels (petrobutanol).

[0010] The chemical synthesis of butanol in one of its isomers can be accomplished by a variety of available methods known in the art (see e.g. Ullmann's Encyclopedia of Industrial Chemistry, 6th edition, 2003, Wiley-VCH Verlag GmbH and Co., Weinheim, Germany, Vol. 5, pp. 716-719). These processes have the disadvantage that they are based on the use of petrochemical derivatives, are generally expensive, and are not environmentally friendly.

[0011] Biological synthesis of butanol can be achieved by fermentation using the acetone-butanol-ethanol (ABE) process carried out by the bacteria *Clostridium acetobutylicum* or other *Clostridium* species. An important disadvantage of the ABE process, however, is that it results in a mixture of acetone, 1-butanol and ethanol. Moreover, the use of bacteria requires sterile process conditions and generally renders the process susceptible to bacteriophage infection. Yeast cells thus provide a very suitable alternative as described above.

SUMMARY OF THE INVENTION

[0012] The present inventors have now identified alternative metabolic routes for increasing the production of cytosolic acetyl-CoA in yeast which can overcome the problems of the PDH by-pass.

[0013] One possible route includes the direct conversion of acetaldehyde to acetyl-CoA without ATP consumption, by use of an acetylating acetaldehyde dehydrogenase (E.C. 1.2.1.10) (see FIG. 2, reaction A, ACDH). Another route includes the direct conversion of pyruvate to acetyl-CoA by an enzyme or a multi-enzyme-complex without ATP consumption, for instance, by use of a pyruvate:NADP oxidoreductase (E.C. 1.2.1.51) see FIG. 2, reaction C, PNO). In these two possible routes, the formation of 1 mole of butanol per mole of glucose would result in the formation of 2 moles of ATP. Yet another route includes the conversion of acetate to acetyl-CoA with 1 ATP consumed per acetyl-CoA formed by an alternative enzyme or a combination of enzymes, for instance, by use of acetate:CoA ligase (ADP-forming, E.C. 6.2.1.13), or by use of ATP:acetate phosphotransferase (E.C. 2.7.2.1) in combination with acetyl-CoA:Pi acetyltransferase (E.C. 2.3.1.8). In this route, the formation of 1 mole of butanol per mole of glucose is ATP-balanced, i.e. no ATP will be formed. The present inventors have now found that such an alternative to the PDH by-pass can result in acetyl-CoA synthesis in the cytosol of the yeast, and that such acetyl-CoA can be used biosynthetically to produce higher amounts of desirable fermentation products, such as butanol.

[0014] In a first aspect, the present invention provides a method of identifying a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) a yeast cell comprising:

[0015] providing a mutated yeast cell, wherein said mutation comprises an inactivation of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS);

[0016] transforming said mutated yeast cell with an expression vector comprising at least one heterologous nucleotide sequence operably linked to a promoter functional in yeast and said at least one heterologous nucleotide sequence encoding at least one candidate polypeptide having potential enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA;

[0017] testing said recombinant mutated yeast cell for its ability to grow on minimal medium containing glucose as sole carbon source, and

[0018] identifying said candidate polypeptide as a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell when growth of said cell is observed.

[0019] In a preferred embodiment of said method the yeast cell is a cell of *Saccharomyces cerevisiae* and the heterologous nucleotide sequence is codon (pair) optimized for expression in *Saccharomyces cerevisiae*.

[0020] In another preferred embodiment, said mutation comprises an inactivation of the gene for acetyl-CoA synthetase isoform 2 (acs2).

[0021] In another preferred embodiment, said at least one candidate polypeptide having enzymatic activity for converting acetaldehyde into acetyl-CoA is a (putative) acetylating acetaldehyde dehydrogenase.

[0022] Alternatively, said at least one heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) a yeast cell may consist of two or more enzymes working

together to achieve the desired conversion from pyruvate, acetaldehyde or acetate into acetyl-CoA.

[0023] In another aspect, the present invention provides an integration vector for the integration in a yeast genome of a heterologous nucleotide sequence encoding a polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA, and the subsequent expression of the heterologous polypeptide therefrom.

[0024] In another aspect, the present invention provides an expression vector expressing heterologous polypeptides in yeast, said expression vector comprising a heterologous nucleotide sequence operably linked to a promoter functional in yeast and said heterologous nucleotide sequence encoding a polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell.

[0025] In a preferred embodiment of said vector the polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA is identified by a method according to the present invention as described above.

[0026] In another preferred embodiment, said polypeptide is selected from SEQ ID NO: 19, 22, 25, 28 and 52 and functional homologues thereof.

[0027] In another preferred embodiment, said expression vector is for expression in *Saccharomyces cerevisiae*, wherein said heterologous nucleotide sequence is codon (pair) optimized for expression in *Saccharomyces cerevisiae*.

[0028] In another preferred embodiment, said heterologous nucleotide sequence is selected from SEQ ID NO: 20, 23, 26 and 29.

[0029] In another aspect, the present invention provides a recombinant yeast cell comprising the expression vector of the present invention as described above.

[0030] In a preferred embodiment, the recombinant yeast cell further comprises an inactivation of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS).

[0031] Preferably, a yeast cell according to the present invention comprises an inactivation of a gene encoding an acetyl-CoA synthetase.

[0032] In another preferred embodiment, the recombinant yeast cell further comprises an inactivation of a gene (nucleotide sequence) encoding an enzyme capable of catalysing the conversion of acetaldehyde to ethanol, preferably a gene encoding an alcohol dehydrogenase.

[0033] As used herein, inactivation of a gene (nucleotide sequence) encoding an enzyme may be achieved by mutation, deletion or disruption of (part of) a gene or nucleotide sequence encoding an enzyme.

[0034] Preferably a yeast cell according to the present invention shows growth on minimal medium containing glucose as sole carbon source.

[0035] In another preferred embodiment of a yeast cell of the invention, said yeast cell further comprises one or more introduced genes encoding a recombinant pathway for the formation of 1-butanol from cytosolic acetyl-CoA. Suitable recombinant pathways from acetyl-CoA to 1-butanol are known in the art. Such pathways are for instance known from WO 2007/041269. Preferably said one or more introduced genes encode enzymes that produce acetoacetyl-CoA, 3-hydroxybutyryl-CoA, crotonyl-CoA, butyryl-CoA, butylaldehyde and/or 1-butanol. Said enzymes can be:

- [0036] acetyl-CoA acetyltransferase (E.C. 2.3.1.9 [Enzyme Nomenclature 1992, Academic Press, San Diego]; although, enzyme's with a broader substrate range (E.C. 2.3.1.16) will be functional as well), which converts 2 moles of acetyl-CoA to acetoacetyl-CoA;
- [0037] NADH-dependent or NADPH-dependent 3-hydroxybutyryl-CoA dehydrogenase E.C. 1.1.1.35 or E.C. 1.1.1.30, resp. E.C. 1.1.1.157 or E.C. 1.1.1.36), which converts acetoacetyl-CoA to 3-hydroxybutyryl-CoA;
- [0038] 3-hydroxybutyryl-CoA dehydratase (also named crotonase; E.C. 4.2.1.17 or E.C. 4.2.1.55), which converts 3-hydroxybutyryl-CoA to crotonyl-CoA;
- [0039] NADH-dependent or NADPH-dependent butyryl-CoA dehydrogenase (E.C. 1.3.1.44 resp. E.C. 1.3.1.38 or E.C. 1.3.99.2), which converts crotonyl-CoA to butyryl-CoA;
- [0040] monofunctional NADH-dependent or NADPH-dependent aldehyde dehydrogenase (E.C. 1.2.1.10, or 1.2.1.57), which converts butyryl-CoA to butyraldehyde, and
- [0041] NADH-dependent or NADPH-dependent butanol dehydrogenase (E.C. 1.1.1.-), which converts butyraldehyde to 1-butanol, or
- [0042] bifunctional NADH-dependent or NADPH-dependent aldehyde/alcohol dehydrogenase (E.C. 1.1.1.1/1.2.1.10), which converts butyryl-CoA to 1-butanol via butyraldehyde
- [0043] In another preferred embodiment of the invention a yeast cell is a *Saccharomyces cerevisiae*.
- [0044] In another aspect, the present invention provides a method of producing butanol, comprising the steps of fermenting a suitable carbon substrate with a yeast cell according to the present invention and recovering the butanol produced during said fermentation.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0045] FIG. 1 is a schematic presentation of the PDH by-pass showing the enzymes pyruvate decarboxylase (PDC; E.C. 4.1.1.1), acetaldehyde dehydrogenase (ALD; E.C. 1.2.1.3, E.C. 1.2.1.4 and E.C. 1.2.1.5), and acetyl-CoA synthetase (ACS; E.C. 6.2.1.1).
- [0046] FIG. 2 shows a schematic metabolic route for butanol production in *Saccharomyces cerevisiae*. Reactions 1-6 are the butanol biosynthesis steps from *Clostridium acetobutylicum* introduced in yeast. A, B, and C indicate alternative reactions for acetyl-CoA biosynthesis in the cytosol. B indicates part of the pyruvate dehydrogenase by-pass (pdc, ald and acs), the natural source of cytosolic acetyl-CoA in yeast. Glc, glucose; EtOH, ethanol; Pyr, Pyruvate; AA, acetaldehyde; ACT, acetate; AcCoA, acetyl-CoA; AACoA, acetoacetyl-CoA; BuCoA, butyryl-CoA; Bual, butylaldehyde; BuOH, butanol; NAD(P)(H), nicotinamide adenine dinucleotide (phosphate) (in reduced form); ATP, adenosine triphosphate; AMP, adenosine monophosphate; TCA cycle, tricarboxylic acid cycle; PDH, pyruvate dehydrogenase; pdc, pyruvate decarboxylase; adh, alcohol dehydrogenase; acdh, acetylating acetaldehyde dehydrogenase; ald, acetaldehyde dehydrogenase; acs, acetyl-CoA synthetase; pno, pyruvate: NADP oxidoreductase. Enzymatic conversions indicated by reaction 1-6 indicate a heterologous butanol pathway from *Clostridium acetobutylicum*: thlB (or ThL) encoding acetyl-CoA acetyltransferase or thiolase [E.C. 2.3.1.9] (SEQ ID NO:30); hbd, 3-hydroxybutyryl-CoA dehydrogenase [E.C. 1.1.1.157] (SEQ ID NO:31); crt, 3-hydroxybutyryl-CoA dehy-

dratase [E.C.4.2.1.55] (SEQ ID NO:32); ter, trans-enoyl CoA reductase; bcd, butyryl-CoA dehydrogenase [E.C.1.3.99.2] (SEQ ID NO:33); etf $\alpha\beta$, heterodimeric electron transfer flavoprotein (etf α and etf β , SEQ ID NO:38 and SEQ ID NO:39, respectively); adhE/adhE1, aldehyde/alcohol dehydrogenase E and E1 [E.C. 1.1.1.1/1.2.1.10] (SEQ ID NO:34 and 35, respectively); bdhA/bdhB, NAD(P)H-dependent butanol dehydrogenase A and B [E.C.:1.1.1.-] (SEQ ID NO:36 and 37, respectively).

[0047] FIG. 3 shows the map of plasmid YEplac112PtdhTadh. The sequence of this plasmid is provided in SEQ ID NO:40.

[0048] FIG. 4 shows an example of a similarity tree based on amino acid sequences of proteins of the types 1 to 4 as described in Example 2 and indicates the branches.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

- [0049] The term "butanol" refers to n-butanol, or 1-butanol.
- [0050] The term "yeast" refers to a phylogenetically diverse group of single-celled fungi, most of which are in the division of Ascomycota and Basidiomycota. The budding yeasts ("true yeasts") are classified in the order Saccharomycetales, with *Saccharomyces cerevisiae* as the most well known species
- [0051] The term "recombinant yeast" as used herein, is defined as a cell which contains a nucleotide sequence and/or protein, or is transformed or genetically modified with a nucleotide sequence that does not naturally occur in the yeast, or it contains additional copy or copies of an endogenous nucleic acid sequence (or protein), or it contains a mutation, deletion or disruption of an endogenous nucleic acid sequence.
- [0052] The term "mutated" as used herein regarding proteins or polypeptides means that at least one amino acid in the wild-type or naturally occurring protein or polypeptide sequence has been replaced with a different amino acid, or deleted from the sequence via mutagenesis of nucleic acids encoding these amino acids. Mutagenesis is a well-known method in the art, and includes, for example, site-directed mutagenesis by means of PCR or via oligonucleotide-mediated mutagenesis as described in Sambrook et al., Molecular Cloning-A Laboratory Manual, 2nd ed., Vol. 1-3 (1989). The term "mutated" as used herein regarding genes means that at least one nucleotide in the nucleotide sequence of that gene or a regulatory sequence thereof, has been replaced with a different nucleotide, or has been deleted from the sequence via mutagenesis, resulting in the transcription of a non-functional protein sequence or the knock-out of that gene.
- [0053] The term "gene", as used herein, refers to a nucleic acid sequence containing a template for a nucleic acid polymerase, in eukaryotes, RNA polymerase II. Genes are transcribed into mRNAs that are then translated into protein.
- [0054] The term pyruvate dehydrogenase (PDH) by-pass refers to the enzymatic cascade from pyruvate to acetyl-CoA in the cytosol of yeast, and which consists of the following enzymes: pyruvate decarboxylase (PDC; E.C. 4.1.1.1) converting pyruvate into acetaldehyde; acetaldehyde dehydrogenase (ALD; E.C. 1.2.1.3, E.C. 1.2.1.4 and E.C. 1.2.1.5), converting acetaldehyde into acetate; and acetyl-CoA synthetase (ACS; E.C. 6.2.1.1), converting acetate into acetyl-CoA.
- [0055] The term "nucleic acid" as used herein, includes reference to a deoxyribonucleotide or ribonucleotide poly-

mer, i.e. a polynucleotide, in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are “polynucleotides” as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including among other things, simple and complex cells.

[0056] The terms “polypeptide”, “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms “polypeptide”, “peptide” and “protein” are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation.

[0057] Sequence identity is herein defined as a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (polynucleotide) sequences, as determined by comparing the sequences. Usually, sequence identities are compared over the whole length of the sequences compared. In the art, “identity” also means the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences.

[0058] Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include BLASTP, BLASTN (Altschul, S. F. et al., *J. Mol. Biol.* 215:403-410 (1990), publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, Md. 20894). Preferred parameters for amino acid sequences comparison using BLASTP are gap open 11.0, gap extend 1, Blosum 62 matrix.

[0059] Every nucleic acid sequence herein that encodes a polypeptide also, by reference to the genetic code, describes every possible silent variation of the nucleic acid. The term “conservatively modified variants” applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to

those nucleic acids which encode identical or conservatively modified variants of the amino acid sequences due to the degeneracy of the genetic code. The term “degeneracy of the genetic code” refers to the fact that a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are “silent variations” and represent one species of conservatively modified variation.

[0060] “Expression” refers to the transcription of a gene into structural RNA (rRNA, tRNA) or messenger RNA (mRNA) with subsequent translation into a protein.

[0061] As used herein, “heterologous” in reference to a nucleic acid or protein is a nucleic acid or protein that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

[0062] As used herein “promoter” is a DNA sequence that directs the transcription of a (structural) gene. Typically, a promoter is located in the 5'-region of a gene, proximal to the transcriptional start site of a (structural) gene. Promoter sequences may be constitutive, inducible or repressible. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent.

[0063] The term “vector” as used herein, includes reference to an autosomal expression vector and to an integration vector used for integration into the chromosome.

[0064] The term “expression vector” refers to a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest under the control of (i.e., operably linked to) additional nucleic acid segments that provide for its transcription. Such additional segments may include promoter and terminator sequences, and may optionally include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, and the like. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both. In particular an expression vector comprises a nucleotide sequence that comprises in the 5' to 3' direction and operably linked: (a) a yeast-recognized transcription and translation initiation region, (b) a coding sequence for a polypeptide of interest, and (c) a yeast-recognized transcription and translation termination region. “Plasmid” refers to autonomously replicating extrachromosomal DNA which is not integrated into a microorganism's genome and is usually circular in nature.

[0065] An “integration vector” refers to a DNA molecule, linear or circular, that can be incorporated in a microorganism's genome and provides for stable inheritance of a gene encoding a polypeptide of interest. The integration vector generally comprises one or more segments comprising a gene sequence encoding a polypeptide of interest under the control of (i.e., operably linked to) additional nucleic acid segments that provide for its transcription. Such additional segments may include promoter and terminator sequences, and one or

more segments that drive the incorporation of the gene of interest into the genome of the target cell, usually by the process of homologous recombination. Typically, the integration vector will be one which can be transferred into the target cell, but which has a replicon which is nonfunctional in that organism. Integration of the segment comprising the gene of interest may be selected if an appropriate marker is included within that segment.

[0066] As used herein, the term “operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence “operably linked” to another control sequence and/or to a coding sequence is ligated in such a way that transcription and/or expression of the coding sequence is achieved under conditions compatible with the control sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

[0067] By “host cell” is meant a cell which contains a vector and supports the replication and/or expression of the vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are cells of the order of Actinomycetales, most preferably yeast cells, most preferably cells of *Saccharomyces cerevisiae*.

[0068] “Transformation” and “transforming”, as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host cell genome.

[0069] The term “oligonucleotide” refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alkyl and aryl-phosphate, phosphorothioate, phosphotriester), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl ribosyl, 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains capable of forming stable secondary structures (e.g., stem-and-loop and loop-stem-loop structures).

[0070] The term “polynucleotide” as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes double- and single-stranded DNA and RNA.

[0071] The term “recombinant polynucleotide” as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; or (2) is linked to a polynucleotide other than that to which it is linked in nature; or (3) does not occur in nature.

[0072] The term “minimal medium” as used herein refers to a chemically defined medium, which includes only the nutrients that are required by the cells to survive and proliferate in culture. Typically, minimal medium is free of biological extracts, e.g., growth factors, serum, pituitary extract, or other

substances, which are not necessary to support the survival and proliferation of a cell population in culture. For example, minimal medium generally includes as essential substances: at least one carbon source, such as glucose; at least one nitrogen source, such as ammonium, ammonium sulfate, ammonium chloride, ammonium nitrate or urea; inorganic salts, such as dipotassium hydrogenphosphate, potassium dihydrogen-phosphate and magnesium sulfate; and other nutrients, such as biotin and vitamins.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0073] A method of the present invention provides a method for identifying heterologous enzymes capable of producing acetyl-CoA in the cytosol of a yeast cell. The heterologous enzyme may produce the acetyl-CoA using pyruvate, acetaldehyde or acetate as a substrate, preferably in a single conversion step. Preferably, the heterologous enzyme produces the acetyl-CoA from acetaldehyde. An enzyme capable of catalyzing said reaction is acetylating acetaldehyde dehydrogenase (acdh; E.C. 1.2.1.10) also referred to as acetaldehyde:NAD⁺ oxidoreductase (CoA-acetylating). The conversion of acetaldehyde into acetyl-CoA by acetylating acetaldehyde dehydrogenase is reversible and runs in the direction of acetyl-CoA when acetaldehyde accumulates in the cytosol. Such an accumulation may for instance be achieved by deletion of alcohol dehydrogenase (adh; E.C. 1.1.1.1).

[0074] The heterologous enzyme may also produce the acetyl-CoA from pyruvate. An enzyme capable of catalyzing said reaction is a pyruvate:NADP oxidoreductase (pno; E.C. 1.2.1.51). The reaction is stoichiometrically identical to the mitochondrial pyruvate dehydrogenase except that pno uses NADPH as a cofactor as compared to PDH that uses NADH. Compared to acdh, an important disadvantage of the pno enzyme system is that pno is oxygen sensitive, and that it is a large multimeric enzyme, and hence, its successful genetic incorporation (a 5-6 kb gene) is much more difficult than that of acdh. For this reason, the use of acdh is preferred in embodiments of the present invention.

[0075] An important feature of a test cell capable of revealing the desired enzymatic activity of a test polypeptide is that the cell is prototrophic as a result of the introduced polypeptide. With this, it is meant that the cell's nutritional requirements do not exceed those of the corresponding wild-type strain and that it will proliferate on minimal medium (in contrast to the auxotroph). In fact, the production of acetyl-CoA as supported by the test polypeptide will cancel the effect of the deletion of said at least one gene of the PDH by-pass, caused by the deletion of the gene for pyruvate decarboxylase (pdc; E.C. 4.1.1.1), acetaldehyde dehydrogenase (aid; E.C. 1.2.1.3, E.C. 1.2.1.4 or E.C. 1.2.1.5), or acetyl-CoA synthetase (acs; E.C. 6.2.1.1). Such complementation assays are well known in the art. In aspects of the present invention the assay is used to identify suitable sources of heterologous enzymes capable of sustaining cytosolic acetyl-CoA production in yeast cells.

[0076] The complementation assay is based on the provision of alternative routes to overcome the deleted enzyme activity of the PDH by-pass. Methods for effecting deletion of genes in yeast are well known in the art, and can for instance be achieved by oligonucleotide-mediated mutagenesis. Good results may be obtained with the plasmid pUG6 carrying the loxP-kanMX-loxP gene disruption cassette (Güldener et al.

[1996] Nucleic Acids Res. 24(13):2519-24; GenPept accession no. P30114). Thus, the skilled person will be able to provide a yeast strain having a deleted acetaldehyde dehydrogenase and/or acetyl-CoA synthetase gene for blocking the PDH by-pass therein.

[0077] *Saccharomyces cerevisiae* comprises two acetyl-CoA synthetase isoforms, Acs1p and Acs2p. Both are the nuclear source of acetyl-CoA for histone acetylation. The production of cytosolic acetyl-CoA is also required for lipid production. Acs activity is essential, since an *acs1 acs2* double null mutant is non-viable. An *acs1* null mutant can grow with ethanol as the sole carbon source. The mutated yeast cell used in aspects of the present invention preferably has an inactivation of the *acs2* gene.

[0078] *Saccharomyces cerevisiae* mutants carrying an inactivation of the *acs2* gene are not able to grow on glucose as sole carbon source, because ACS1 is repressed and the protein is actively degraded. Complementation of such a delta *acs2* mutant with a plasmid based *acs* gene will restore the cell's ability to grow on glucose as single carbon source. In addition, growth of such a mutant is complemented by the expression of genes supporting alternative routes for the production of sufficient cytosolic acetyl-CoA. Thus, transformation of the delta *acs2* mutant with a plasmid from which a functional (heterologous) *acdh* or *pno* can be expressed will restore the mutant's ability to grow on glucose as sole carbon source. It should be understood that in addition to the removal of the ACS2 locus, one may also remove the ACS1 locus. Although it is believed that this may in some instances prevent the occurrence of revertants (mutations in the ACS1 locus leading to reversion of the delta *acs2* phenotype), this was however not found to be essential. Double mutants (*acs1/acs2Δ* strains) would be wholly dependant on the introduced *acdh* or *pno* gene for the production of cytosolic acetyl-CoA.

[0079] An important advantage of a complementation assay of the present invention is that it can be performed as a plate screening assay wherein successful complementation is observed as colony growth. This is much faster than experiments that require the analysis for the production of a desired metabolic product.

[0080] For complementation of the mutation, the yeast cell having the inactivated *ald* and/or *acs* gene is then transformed with a suitable expression vector comprising a nucleotide sequence of a heterologous test polypeptide.

[0081] Yeast expression vectors are widely available from a variety of commercial suppliers. To date, functional complementation of yeast mutations by foreign homologues has become a standard practice in engineering of *Saccharomyces cerevisiae*. Suitable expression vectors for heterologous gene expression may be based on artificial, inducible promoters such as the GAL promoter, but is preferably based on constitutive promoters such as the TDH3 promoter. Suitable systems are exemplified in the examples below. In certain production systems, the use of an inducible promoter may be preferred, as it would allow for temporal separation of stages for biomass production (promotor not induced) and fermentation product production (promotor induced). In another highly preferred embodiment in certain production systems, the vector is in integration vector for stable integrating the heterologous genes in the genome of the yeast production strain.

[0082] In order to achieve optimal expression in yeast, the codon (pair) usage of the heterologous gene may be optimized by using any one of a variety of synthetic gene design

software packages, for instance GeneOptimizer® from Geneart AG (Regensburg, Germany) for codon usage optimization or codon pair usage optimization as described in WO2008/000632. Such adaptation of codon usage ensures that the heterologous genes, which are for instance of bacterial origin, are effectively processed by the yeast transcription and translation machinery. Optimization of codon pair usage will result in enhanced protein expression in the yeast cell.

[0083] The optimized sequences may for instance be cloned into a high copy yeast expression plasmid, operably linked to a (preferably constitutive) promoter functional in yeast. Good results have been obtained with the plasmid YEplac112 (2μ TRP1) (Gietz & Sugino [1988] Gene 74(2): 527-34).

[0084] Heterologous genes that encode a candidate polypeptide having potential enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA may be identified in silico. Suitable enzymes described as possessing the capacity to convert acetaldehyde into acetyl-CoA are acetylating acetaldehyde dehydrogenases (E.C. 1.2.1.10). The nucleotide and amino acid sequences of over 200 of these enzymes from a variety of microbial origins are described in various databases (e.g. the KEGG (Kyoto Encyclopedia of Genes and Genomes) database).

[0085] The present inventors have selected several acetylating acetaldehyde dehydrogenases and tested these in the delta *acs2* mutant-based assay system of the present invention. Many of these, though not all, were functional in *S. cerevisiae* when codon pair usage was optimized.

[0086] Functional homologues to these proteins can also be used in aspects of the present invention. The term "functional homologues" as used herein refers to a protein comprising the amino acid sequence of SEQ ID NO: 19, 22, 25 or the acetaldehyde dehydrogenase part of SEQ ID NOs: 28 and 52 in which one or more amino acids are substituted, deleted, added, and/or inserted, and which protein has the same enzymatic functionality for substrate conversion, for instance an acetylating acetaldehyde dehydrogenase homologue is capable of converting acetaldehyde into acetyl-CoA. This functionality may be tested by use of an assay system comprising a recombinant yeast cell comprising an expression vector for the expression of the homologue in yeast, said expression vector comprising a heterologous nucleotide sequence operably linked to a promoter functional in yeast and said heterologous nucleotide sequence encoding the homologous polypeptide of which enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl CoA in (the cytosol of) said yeast cell is to be tested, and performing a method for identifying a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) a yeast cell as described herein using said assay system. Candidate homologues may be identified by using in silico similarity analyses. A detailed example of such an analysis is described in Example 2 below. The skilled person will be able to derive therefrom how suitable candidate homologues may be found and, optionally upon codon (pair) optimization, will be able to test the required functionality of such candidate homologues using the assay system of the present invention as described above. A suitable homologue represents a polypeptide having an amino acid sequence identity to an acetylating acetaldehyde dehydrogenase of more than 50%, preferably more than 60%, more preferably more than 70%, 80%, 90% or more, for instance having such an amino acid sequence

identity to SEQ ID NOs:19, 22, 25, or the acetaldehyde dehydrogenase part of SEQ ID NOs:28 and 52 and having the required enzymatic functionality for converting acetaldehyde into acetyl-CoA. Similarly, enzymes described for the direct conversion of pyruvate into acetyl-CoA and the functional homologues thereof, as well as enzymes described for the conversion of acetate to acetyl-CoA and the functional homologues thereof, can also be used, similar as described for acetylating acetaldehyde dehydrogenase above.

[0087] A method of the present invention further comprises the step of testing the ability of the mutated and test-protein transformed yeast cell to grow on minimal medium containing glucose as sole carbon source. As stated earlier, this may suitably occur on solid (agar) media in Petri dishes (plates) where growth can be observed as growth of a colony, however, liquid media are equally suitable and growth may be detected by turbidity. Other methods for determining growth of the mutated and test-protein transformed yeast cell on minimal medium containing glucose as sole carbon source may also be used.

[0088] When the mutated and test-protein-transformed yeast cell is capable of growth on minimal medium with glucose, the candidate polypeptide is successfully identified as a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell. Growth may suitably be observed as colony formation on solid growth media, in particular minimal medium containing glucose.

[0089] An expression vector for the expression of heterologous polypeptides in yeast, according to the present invention may be any expression vector suitable for transforming yeast. Innumerable examples are available in the art that can suitably be used to express heterologous nucleotide sequences in yeast. A very suitable vector in aspects of the invention is a plasmid. A highly preferred plasmid is YEplac112PtdhTadh (SEQ ID NO:40).

[0090] Generally, the heterologous nucleotide sequence encoding the polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl CoA in (the cytosol of) said yeast cell, will be placed under control of a promoter functional in yeast. Preferably the promoter is a constitutive promoter. The promoter on plasmid YEplac112PtdhTadh is the TDH3 promoter.

[0091] The heterologous nucleotide sequences incorporated in the expression vector of the present invention may be any pno, acdh or other enzyme capable of converting pyruvate, acetaldehyde or acetate (respectively) into acetyl-CoA in the cytosol of the yeast. Preferred nucleotide sequences are those as identified herein, namely the nucleotide sequences encoding:

[0092] the ethanolamine utilization protein EutE from *E. coli* HS (nucleotide sequences with SEQ ID NO:18);

[0093] the hypothetical protein Lin1129 from *Listeria innocua* similar to ethanolamine utilization protein EutE, (nucleotide sequences with SEQ ID NO:21)

[0094] the acetaldehyde dehydrogenase EDK33116 from *Clostridium kluyveri* DSM 555 (nucleotide sequences with SEQ ID NO:24); and

[0095] the adhE homologue of *S. aureus* (nucleotide sequences with SEQ ID NO:27) encoding a bifunctional acetaldehyde/alcohol dehydrogenase in *Staphylococcus aureus* subsp. *aureus* N315, or the acetaldehyde dehydrogenase functional part thereof.

[0096] the adhE homologue of *Piromyces* sp. E2 (nucleotide sequence SEQ ID NO: 51) encoding a bifunctional acetaldehyde/alcohol dehydrogenase, or the acetaldehyde dehydrogenase part thereof.

[0097] Also suitable are functional homologues of these nucleotide sequences, or of the polypeptides that they encode. With this term is meant that a nucleic acid sequence having more than 80%, 90% or 95% sequence identity with the nucleotide sequences encoding the above acdh enzymes, or having more than 50%, preferably more than 60%, 70%, 80%, 90%, or 95% sequence identity with the amino acid sequence of the above acdh enzymes, with the proviso that the polypeptides encoded by the homologous sequences exhibit functional enzymatic acdh activity.

[0098] As stated above, these nucleotide sequences can be optimized for expression in *Saccharomyces cerevisiae* by optimization of codon pair usage well known in the art. Codon pair optimized sequences for the SEQ ID NO:18, 21, 24, and 27 are provided in SEQ ID NO:20, 23, 26, and 29, respectively.

[0099] The expression vector of the invention may be used to transform a yeast cell. Methods of transformation include electroporation, glass bead and biolistic transformation, all of which are well known in the art and for instance described in Sambrook et al., Molecular Cloning-A Laboratory Manual, 2nd ed., Vol. 1-3 (1989).

[0100] A yeast cell according to the present invention comprises a heterologous nucleotide sequence encoding a polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell. Preferably, a yeast cell of the invention comprises a heterologous acdh or pno. The advantage of such a yeast cell is that it can produce acetyl-CoA by a metabolic route wherein the PDH by-pass is not required. This is energetically more favourable under anaerobic conditions, and may form the basis of any biological synthesis process using yeast cells under anaerobic conditions where acetyl-CoA is an intermediate. In addition to comprising the heterologous acdh or pno, the yeast cell of the invention may comprise various gene deletions or gene supplementations, depending on the intended use of the yeast.

[0101] Preferably a yeast cell according to the present invention comprises an inactivation of a nucleotide sequence (gene) encoding an enzyme capable of catalysing the conversion of acetaldehyde to ethanol, preferably an alcohol dehydrogenase, for instance to optimize acetaldehyde accumulation in the yeast cell.

[0102] If used in a method of screening for heterologous enzymes according to a method of the invention, the yeast cell comprises a deletion of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS), preferably acetyl-CoA synthetase, most preferably acs2.

[0103] If used in a method of producing a fermentation product, the yeast cell may optionally comprise a number of (heterologous) gene supplementations supporting the metabolic pathway from acetyl-CoA to said butanol. Such a pathway may consist only of heterologous gene products, or may make use of a mixture of heterologous and endogenous gene products. In the event the fermentation product is butanol, use can be made of a yeast comprising genes encoding enzymes for the butanol pathway of e.g. *Clostridium acetobutylicum* as described herein and in FIG. 2. In the event the yeast cell

according to the present invention comprises genes encoding enzymes for butanol production, the yeast preferably comprises a nucleotide sequence encoding a butyryl-CoA dehydrogenase and at least one nucleotide sequence encoding a heterologous electron transfer flavoprotein (ETF). It was found that a yeast cell comprising an ETF in addition to genes of the butanol pathway produces an increased amount of butanol.

[0104] A heterologous electron transfer flavoprotein in the eukaryotic cell according to the present invention may be a single protein or the ETF may comprise two or more subunits, for instance an alpha and a beta subunit. Preferably the ETF comprises an ETF alpha (SEQ ID NO: 38) and an ETF beta (SEQ ID NO: 39). The electron transfer flavoprotein may be derived from any suitable origin. Preferably, the ETF is derived from the same origin as the butyryl-CoA dehydrogenase. Preferably, the ETF is derived from prokaryotic origin preferably from a *Clostridium* sp., preferably a *Clostridium acetobutylicum* or a *Clostridium beijerinckii*.

[0105] A method for producing a fermentation product according to the present invention, preferably comprises growing a yeast under anaerobic conditions on a suitable carbon and energy source. Suitable sources of carbon and energy are C5 and C6 sugars (monosaccharides) such as glucose and polysaccharides such as starch. Other raw materials such as sugarcane, maize, wheat, barley, sugarbeets, rapeseed, and sunflower are also suitable. In some instances the raw material may be pre-digested by enzymatic treatment. Most preferably the carbon source is lignocellulose, which is composed of mainly cellulose, hemicellulose, pectin, and lignin. Lignocellulose is found, for example, in the stems, leaves, hulls, husks, and cobs of plants. Hydrolysis of these polymers by specific enzymatic treatment releases a mixture of neutral sugars including glucose, xylose, mannose, galactose, and arabinose. Lignocellulosic materials, such as wood, herbaceous material, agricultural residues, corn fiber, waste paper, pulp and paper mill residues can be used to produce butanol. Hydrolysing enzymes are for instance beta-linked glucans for the hydrolysis of cellulose (these enzymes include endoglucanases, cellobiohydrolases, glucohydrolases and beta-glucosidases); beta-glucosidases hydrolyze cellobiose; endo-acting and exo-acting hemicellulases and cellobiases for hydrolysis of hemicellulose, and acetylsterases and esterases that hydrolyze lignin glycoside bonds. These and other methods for hydrolysis of lignocellulose are well known in the art.

[0106] Variations and modifications of the embodiments disclosed herein are possible, and practical alternatives to and equivalents of the various elements of the embodiments would be understood to those of ordinary skill in the art upon study of this patent document. These and other variations and modifications of the embodiments disclosed herein may be made without departing from the scope and spirit of the invention.

[0107] The invention will now be illustrated by way of the following non-limiting examples.

EXAMPLES

[0108] The following examples illustrate the provision of a strain of *Saccharomyces cerevisiae* useful in assays and methods of the present invention, for instance in methods for identifying heterologous enzymes capable of forming cytosolic acetyl-CoA in *S. cerevisiae*. Such methods are useful in

the identification of routes/enzymes which allow the cytosolic supply of acetyl-CoA in *S. cerevisiae* under anaerobic conditions.

[0109] In order to enhance cytosolic acetyl-CoA formation in our butanol production strain, a selection method was set up to identify heterologous enzymes forming cytosolic acetyl-CoA in *S. cerevisiae*. The test system is based on a delta *acs2* yeast mutant deficient in cytosolic acetyl-CoA biosynthesis on glucose, such a strain is unable to grow on glucose as sole carbon source unless cytosolic acetyl-CoA formation is complemented. Complementation studies in such a strain can reveal which heterologous enzymes are suitable for use in butanol producing strains of *Saccharomyces cerevisiae*.

[0110] Acetylating acetaldehyde dehydrogenase was identified to be a good candidate for cytosolic acetyl-CoA supply over the homologous PDH by-pass because no ATP is dissipated. Twelve putative acetylating acetaldehyde dehydrogenases, identified based on sequence homology, were synthesized and checked for complementation of the delta *acs2* yeast.

[0111] The codon pair optimized genes of the *eutE* homologues of *E. coli*, *L. innocua* and *C. kluyveri* and the *adhE* homologue of *S. aureus* were able to complement the *acs2* yeast mutants (4 out of 7), resulting in growth of the *acs2Δ S. cerevisiae* host. The aim is to improve butanol biosynthesis in yeast by expression of one or more genes so identified.

[0112] In order to test if these heterologous routes for cytosolic acetyl-CoA supply work in *S. cerevisiae*, a screening system was developed based on *Saccharomyces cerevisiae* mutants carrying a deletion of the *acs2* gene. These cells are not able to grow on glucose as sole carbon source unless the delta *acs2* mutant is complemented with a plasmid based *acs* gene or complemented with the expression of any other gene generating sufficient cytosolic acetyl-CoA. So if it were to be transformed with a plasmid leading to active expression of *acdh* or *pno*, such a mutant should be able to grow again with glucose as single carbon source. The complementation studies were performed on plates. The following experiments were performed to set up and evaluate the test system.

Example 1

Construction of Delta *acs2* Strain

[0113] The *S. cerevisiae acs2* deleted strain (*acs2Δ* strain) was produced by first performing a PCR on plasmid pUG6 (Güldener et al., 1996, supra) with the following oligonucleotides:

```
5' acs2 :: Kanlox
5' -tacacaaacagaatacaggaaagtaaatcaatacaataataaacag
ctgaagcttcgtacgc-3'

3' acs2 :: Kanlox
5' -tctcattacgaaatctttctcatttaagttatttcttttttgaggc
ataggccactagtgatctg-3'.
```

[0114] The resulting 1.4 kb fragment, containing the KanMX marker which confers resistance to G418, was used to transform *S. cerevisiae* CEN.PK113-3C (MATA *trp1-289*). After transformation the strain was plated on YPD (10 g l⁻¹ yeast extract (BD Difco), 20 g l⁻¹ peptone (BD Difco)), 10 g

1⁻¹ glucose) with 200 mg/ml Geneticin (G418). In resistant transformants, correct integration was verified by PCR using oligonucleotides:

5' ACS2:	5'-gatattcggtagccgattcc-3'
3' ACS2:	5'-ccgtaaaccttctcgtaatgc-3'
ACS2internal:	5'-cggattcgatcatcagcttca-3'
KanA:	5'-cgcacgtcaagactgtcaag-3'
KanB:	5'-tcgtatgtgaatgctggtcg-3'

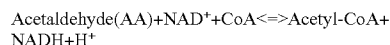
[0115] The phenotype was verified by testing for growth on YP with 1% glucose (YPD) or 1% ethanol+1% glycerol (YPEG) as the carbon source.

[0116] One transformant that had the correct PCR bands and did not grow on YP with glucose, but did grow on with YP with ethanol and glycerol as the carbon sources, was picked and named RWB060 (MATA trp1-289 acs2::Kanlox).

Example 2

In Silico Identification of Putative Acetylating Acetaldehyde Dehydrogenases for Direct Conversion of Acetaldehyde to Acetyl-CoA

[0117] Enzymes described for the conversion of acetaldehyde to acetyl-CoA are the so-called acetylating acetaldehyde dehydrogenases (ACDH) (E.C. 1.2.1.10) catalysing the following reaction:



[0118] From literature four types of proteins have been described that have this activity:

[0119] 1) Bifunctional proteins that catalyze the reversible conversion of acetyl-CoA to acetaldehyde, and the subsequent reversible conversion of acetaldehyde to ethanol. An example of this type of proteins is the AdhE protein in *E. coli* (GenBank No: NP_415757). AdhE appears to be the evolutionary product of a gene fusion. The NH₂-terminal region of the AdhE protein is highly homologous to aldehyde:NAD⁺ oxidoreductases, whereas the COOH-terminal region is homologous to a family of Fe²⁺ dependent ethanol:NAD⁺ oxidoreductases (Membrillo-Hernandez et al., (2000) J. Biol. Chem. 275: 33869-33875). The *E. coli* AdhE is subject to metal-catalyzed oxidation and therefore oxygen-sensitive (Tamarit et al. (1998) J. Biol. Chem. 273:3027-32).

[0120] 2) Proteins that catalyze the reversible conversion of acetyl-CoA to acetaldehyde in strictly or facultative anaerobic micro-organisms but do not possess alcohol dehydrogenase activity. An example of this type of proteins has been reported in *Clostridium kluyveri* (Smith et al. (1980) Arch. Biochem. Biophys. 203: 663-675). An acetylating acetaldehyde dehydrogenase has been annotated in the genome of *Clostridium kluyveri* DSM 555 (GenBank No: EDK33116). A homologous protein AcdH is identified in the genome of *Lactobacillus plantarum* (GenBank No: NP_784141). Another example of this type of proteins is the ald gene product in *Clostridium beijerinckii* NRRL B593 (Toth et al. (1999) Appl. Environ. Microbiol. 65: 4973-4980, GenBank No: AAD31841).

[0121] 3) Proteins that are involved in ethanolamine catabolism. Ethanolamine can be utilized both as carbon and nitrogen source by many enterobacteria (Stojiljkovic et al.

(1995) J. Bacteriol. 177: 1357-1366). Ethanolamine is first converted by ethanolamine ammonia lyase to ammonia and acetaldehyde, subsequently, acetaldehyde is converted by acetylating acetaldehyde dehydrogenase to acetyl-CoA. An example of this type of acetylating acetaldehyde dehydrogenase is the EutE protein in *Salmonella typhimurium* (Stojiljkovic et al. (1995) J. Bacteriol. 177: 1357-1366, GenBank No: AAL21357). *E. coli* is also able to utilize ethanolamine (Scarlett et al. (1976) J. Gen. Microbiol. 95:173-176) and has an EutE protein (GenBank No: AAG57564) which is homologous to the EutE protein in *S. typhimurium*.

[0122] 4) Proteins that are part of a bifunctional aldolase-dehydrogenase complex involved in 4-hydroxy-2-ketovalerate catabolism. Such bifunctional enzymes catalyze the final two steps of the meta-cleavage pathway for catechol, an intermediate in many bacterial species in the degradation of phenols, toluates, naphthalene, biphenyls and other aromatic compounds (Powlowski and Shingler (1994) Biodegradation 5, 219-236). 4-Hydroxy-2-ketovalerate is first converted by 4-hydroxy-2-ketovalerate aldolase to pyruvate and acetaldehyde, subsequently acetaldehyde is converted by acetylating acetaldehyde dehydrogenase to acetyl-CoA. An example of this type of acetylating acetaldehyde dehydrogenase is the DmpF protein in *Pseudomonas* sp CF600 (GenBank No: CAA43226) (Shingler et al. (1992) J. Bacteriol. 174:711-24). *E. coli* has a homologous MphF protein (Ferrández et al. (1997) J. Bacteriol. 179: 2573-2581, GenBank No: NP_414885) to the DmpF protein in *Pseudomonas* sp. CF600.

[0123] To identify the protein family members of acetylating acetaldehyde dehydrogenase, the amino acid sequences of the *E. coli* bifunctional AdhE protein (GenBank No: NP_415757), *L. plantarum* AcdH protein (acetylating) (GenBank No: NP_784141), the *E. coli* EutE protein (GenBank No: AAG57564) and the *E. coli* MhpF protein (GenBank No: NP_414885) were each run as a query sequence in a BLASTp search against the GenBank non-redundant protein database using default parameters. Amino acid sequences with an E-value smaller or equal to 1 e-20 were extracted. Redundant sequences were removed and the remaining sequences were aligned and a similarity tree was built using Genedata Physolopher protein analyzer software, version 6.5. 2. A similarity tree provides information on organism sequence similarity. The tree is created independently of the ClustalW algorithm by pairwise comparison of the amino acid sequences per residue position. At each position, the similarity is rated and summed up to an overall score for each sequence pair. Based on these pairwise scores a hierarchical clustering is performed, which arranges the sequences in a tree. Note that the ald gene product of *C. beijerinckii* (GenBank no: AAD31841) clustered together with the EutE proteins from *E. coli* and *S. typhimurium*. From this similarity tree four major branches could be defined, each branch contains one amino acid sequence that was used as a query for the BLASTp search. FIG. 4 shows an example of such a similarity tree, containing all sequences that are mentioned in this example.

[0124] At least one amino acid sequence was selected from each branch for complementation tests in *S. cerevisiae* delta acs2. Preferably, the selected amino acid sequences have experimental evidence of its biochemical function as acety-

lating acetaldehyde dehydrogenase. Such evidences can be found in public databases, such as in the BRENDA, UniProt and NCBI Entrez databases.

Example 3

Construction of Expression Plasmids and Complementation Test

[0125] To test whether acetylating acetaldehyde dehydrogenases (ACDH) could complement the deletion of ACS2 in *S. cerevisiae*, several genes coding for a (putative) ACDH were chosen from a variety of databases as described above.

[0126] To achieve optimal expression in yeast, the codon usage of all genes was adapted by codon pair optimization. These sequences were synthesized at Genart AG (Regensburg, Germany).

[0127] The optimized sequences were cloned into the high copy yeast expression plasmid YEplac112PtdhTadh (SEQ ID NO:40; based on YEplac112 (2 μ TRP1) (Gietz & Sugino [1988] Gene 74(2):527-34), allowing constitutive expression from the TDH3 promoter.

[0128] YEplac112PtdhTadh was made by cloning a KpnI-SacI fragment from p426GPD (Mumberg et al. [1995] Gene. 156(1):119-22), containing the TDH3 promoter and CYC1 terminator, into YEplac112 cut with KpnI-SacI. The resulting plasmid was cut with KpnI and SphI and the ends were made blunt then ligated to give YEplac112TDH. To obtain YEplac112PtdhTadh, YEplac112TDH was cut with PstI-HindIII and ligated to a 345 by PstI-HindIII PCR fragment containing the ADH1 terminator (Tadh), thus replacing the CYC1 terminator and changing the polylinker between the promoter and terminator. The Tadh PCR fragment was generated using the following oligonucleotides:

MCS-5' Tadh:

5' - aaggtagctctagactagtcctccgggctgcagtcgactcgagcgaatt

tcttatgatttatgatt-3'

Tadh1-Hind:

5' - aggaagcttaggcctgtgtggaagaacgattacaacagg-3'

[0129] PCR was done with Vent^R DNA polymerase, according to the manufacturer's specifications.

[0130] The synthetic constructs containing the ACDH genes were cut with SpeI-PstI and ligated into YEplac112PtdhTadh digested with the same enzymes, resulting in pBOL058 through to pBOL068 and pBOL082. The names of the final plasmids and the genes they contain are given in Table 1.

TABLE 1

Organisms	Name	Group*	Size (kb)	SEQ ID NO. DNA/PRT/OPT
<i>Escherichia coli</i>	adhE	1	2.6	
<i>Entamoeba histolytica</i>	adh2	1	2.6	48/50/49
<i>Staphylococcus aureus</i>	adhE	1	2.6	27/28/29
<i>Piromyces</i> sp. E2	adhE	1	2.6	51/52

TABLE 1-continued

Table 1: Overview on putative acetylating acetaldehyde dehydrogenases tested for complementation of delta acs2 *S. cerevisiae* strain. Genes which resulted in complementation are given in bold. SEQ ID NOs are provided for the DNA sequence of the wild type gene, the protein expressed therefrom, and the codon pair optimized DNA sequence.

Organisms	Name	Group*	Size (kb)	SEQ ID NO. DNA/PRT/OPT
<i>Clostridium kluyveri</i>	EDK33116	2	1.5	24/25/26
<i>Lactobacillus plantarum</i>	acdH	2	1.4	
<i>Escherichia coli</i>	EutE	3	1.4	18/19/20
<i>Listeria innocua</i>	Lin1129	3	1.4	21/22/23
<i>Pseudomonas putida</i>	YP 001268189	4	1.0	

*Group refers to the group of proteins having ACDH activity as defined in Example 2. Group 1: similar to bifunctional *E. coli* AdhE (AdhE-type of proteins); group 2: proteins having similarity to *Lactobacillus plantarum* AcdH (AcdH-type of proteins); group 3: similar to *E. coli* EutE (EutE-type of proteins); group 4: similar to *E. coli* MhpF (MhpF-type of proteins).

[0131] All plasmids were used to transform the delta acs2 yeast strain RWB060. As negative control, the empty vector YEplac112 was used. Transformants were plated on mineral medium (Verduyn et al. [1992] Yeast 8 (1992), pp. 501-517) containing either 1% glucose (MYD) or 1% ethanol+1% glycerol (MYEG) as single carbon source.

[0132] While for all constructs several transformants could be selected on minimal medium with ethanol/glycerol, this was not the case on the glucose containing plates.

TABLE 2

Result of a complementation experiment for putative acetylating acetaldehyde dehydrogenases in delta acs2 *S. cerevisiae* strain RWB060. Genes resulting in complementation are given in bold. MYEG and MYD columns indicate number of transformants on plates MYEG (ethanol/glycerol) and MYG (glucose).

Organisms	Gene (GenPept accession)	plasmid	MYEG	MYD
	none	YEplac112	75	0
<i>Escherichia coli</i>	adhE	pBOL059	6	0
<i>Entamoeba histolytica</i>	adh2	pBOL061	54	0
<i>Staphylococcus aureus</i>	adhE	pBOL064	36	39
	(BAB41363)			
<i>Piromyces</i> sp. E2	adhE	pBOL139	32	3
<i>Clostridium kluyveri</i>	EDK33116	pBOL065	21	8
	(EDK33116)			
<i>Lactobacillus plantarum</i>	acdH	pBOL058	6	0
<i>Escherichia coli</i>	EutE	pBOL066	24	18
	(ABV06849)			
<i>Listeria innocua</i>	Lin1129	pBOL067	28	8
	(CAC96360)			
<i>Pseudomonas putida</i>	YP 001268189	pBOL068	32	0

[0133] On the glucose containing plates, transformants could only be selected for plasmids pBOL064, pBOL065, pBOL066, and pBOL067, not the empty vector. There was also a clear difference in colony size, depending on the plasmid used. While construct pBOL066 (*E. coli* eutE) resulted in biggest colonies, colonies of pBOL067 (*L. innocua* lin1129) appeared a bit smaller and pBOL065 (*C. kluyveri* edk3116) showed smallest colonies. Plasmid pBOL064 (*S. aureus* adhE) and plasmid pBOL139 (*Piromyces* sp. E2, adhE) were done at a later date, so could not be compared directly. Colonies containing pBOL64 seemed to be similar to colonies comprising pBOL066 and colonies comprising pBOL139 seemed to be similar to colonies comprising pBOL065.

[0134] To ensure that these results did not arise from spontaneous revertants, transformation experiments were repeated for some of the plasmids, giving the same results. In addition, for almost all plasmids four transformants were selected at random from the MYEG plates and restreaked onto MYD and MYEG plates.

[0135] In all experiments no growth was ever seen on glucose with the empty vector (YEplac112), while only pBOL065, pBOL066 and pBOL067 repeatedly gave good growth on glucose. Plasmid pBOL064 was not re-tested this way after the initial very positive result.

[0136] From these results, it was concluded that the codon pair optimized genes of the eutE homologues of:

[0137] *E. coli* (SEQ ID NO:20) encoding the ethanolamine utilization protein EutE from *E. coli* HS;

[0138] *L. innocua* (SEQ ID NO:23) encoding a hypothetical protein from *L. innocua* similar to ethanolamine utilization protein EutE, and

[0139] *C. kluyveri* (SEQ ID NO:26) encoding acetylating acetaldehyde dehydrogenase in *Clostridium kluyveri* DSM 555;

and the codon pair optimized gene of the adhE homologue of

[0140] *S. aureus* (SEQ ID NO:29) encoding a bifunctional acetaldehyde/alcohol dehydrogenase in *Staphylococcus aureus* subsp. *aureus* N315;

and the non codon pair optimized gene of the adhE homologue

[0141] *Piromyces* sp. E2 (SEQ ID NO:51) encoding a bifunctional acetaldehyde/alcohol dehydrogenase

are able to complement the acs2 yeast mutants. These genes encode an enzymatic activity allowing the formation of cytosolic acetyl-CoA from acetaldehyde in yeast.

Conclusions

[0142] The supply of cytosolic acetyl-CoA is believed to be a bottleneck in the butanol production in yeast. In order to identify heterologous genes encoding for enzymes forming cytosolic acetyl-CoA in *S. cerevisiae* a test system based on a delta acs2 yeast mutant was established.

[0143] Due to its deficiency in cytosolic acetyl-CoA biosynthesis on glucose, the acs2Δ strain is unable to grow with glucose as sole carbon source.

[0144] 9 putative acetylating acetaldehyde dehydrogenases identified as candidates for cytosolic acetyl-CoA supply from acetaldehyde were expressed in the acs2Δ yeast. In total, 5 of these 9 genes complemented growth of the acs2Δ strain with glucose as single carbon source. Therewith, the use of the delta acs2 strain as pre-selection tool for feasible routes for cytosolic supply of acetyl-CoA was shown.

[0145] 4 of 5 acetylating acetaldehyde dehydrogenases identified thus far, eutE homologues of *E. coli*, *L. innocua* and *C. kluyveri* and the adhE homologue of *S. aureus*, and *Piromyces* sp. E2, were successfully integrated in butanol producing strains of *S. cerevisiae*. The effect on butanol production was investigated as described in Examples below.

[0146] This test system may also be used, to analyse whether pyruvate:NADP oxidoreductase can successfully be over-expressed in yeast. Due to the oxygen sensitivity, this test has to be performed anaerobically.

[0147] Examples 4-6 below describe the testing 4 of the 5 selected ACDH genes from Example 3 for improvement of butanol production.

Example 4

Construction of a Butanol Producing Yeast Strain and Knocking Out the ADH1 and ADH2 Genes

[0148] The six *Clostridium acetobutylicum* genes involved in butanol biosynthesis from Acetyl-CoA are listed in Table 3. The genes were codon pair optimized for *S. cerevisiae* as described in WO2008/000632 and expressed from yeast promoters and terminators as listed in Table 3.

[0149] Two yeast integration vectors (pBOL34 [SEQ ID NO:41] and pBOL36 [SEQ ID NO:42]), each containing 3 of the six codon pair optimised genes from *Clostridium acetobutylicum* involved in butanol biosynthesis, were designed and synthesized at Geneart.

[0150] The genes ThiL, Hbd and Crt are expressed from pBOL34 containing a AmdS selection marker. The final three genes, Bcd, BdhB and AdhE were expressed from a integration vector with an AmdS selection marker named pBOL36.

TABLE 3

Genes used for butanol production in <i>S. cerevisiae</i> including the promoter (1000 bp) and terminator (500 bp)			
Gene	activity	Promotor	Terminator
ThiL	acetyl CoA c-acetyltransferase [E.C. 2.3.1.9]	ADH1	TDH1
Hbd	3-hydroxybutyryl-CoA dehydrogenase [E.C.1.1.1.157]	ENO1	PMA1
Crt	3-hydroxybutyryl-CoA dehydratase [E.C.4.2.1.55]	TDH1	ADH1
Bcd	butyryl-CoA dehydrogenase [E.C.1.3.99.2]	PDC1	TDH1
BdhB	NADH-dependent butanol dehydrogenase [E.C.1.1.1.—]	ENO1	PMA1
adhE	alcohol/acetaldehyde CoA dehydrogenase [E.C.: 1.1.1.1/1.2.1.10]	TDH1	ADH2

[0151] For integration in the ADH2 locus, pBOL36 was linearized by a BsaBI digestion. *S. cerevisiae* CEN.PK113-5D (MATa MAL2-8c SUC2 ura3-52) was transformed with the linear fragment and grown on plates with YCB (Difco) and 5 mM acetamide as nitrogen source.

[0152] The AmdS marker was removed by recombination by growing the transformants for 6 hours in YEPD in 2 ml tubes at 30° C. Cells were subsequently plated on 1.8% agar medium containing YCB (Difco) and 40 mM fluoracetamide and 30 mM phosphate buffer pH 6.8 supporting growth only from cells that have lost the AmdS marker. Correct integration and recombination were confirmed by PCR. The correct integration of the fragment upstream was confirmed with the following primers:

P1 : 5' -GAATTGAAGGATATCTACATCAAG-3' and

P2 : 5' -CCCATCTACGGAACCTGATCAAGC-3'.

[0153] The correct integration of the fragment downstream was confirmed with the following primers:

P3 : 5' -GATGGTGTACCATACCAGGCTAG-3' and

P4 : 5' -GTTCTCTGGTCAAGTTGAAGTCCATTTTGATTGATTGACTGTGTTA

TTTTGCGTG-3'.

[0154] The resulting strain was named BLT021.

[0155] pBOL34 was linearized by a PstI digestion and integrated in the ADH1 locus of BLT021. The transformants were grown on plates containing YCB (Difco) and 5 mM acetamide. For removal of the AmdS selection marker, colonies were inoculated in YEPD and grown for 6 hours in 2 ml tubes at 30° C. The cells were plated on YCB (Difco) and 40 mM fluoracetamide and 0.1% ammonium sulphate.

[0156] Correct integration and recombination were confirmed by PCR. The correct integration of the fragment upstream was confirmed with the following primer set:

P5: 5' -GAACAATAGAGCGACCATGACCTTG -3' and
P6: 5' -GACATCAGCGTCACCAGCCTTGATG -3'.

[0157] The correct integration of the fragment downstream was confirmed with the following primer set:

P7: 5' -GATTGAAGGTTTCAAGAACAGGTGATG -3' and
P8: 5' -GGCGATCAGAGTTGAAAAAAATG -3'.

[0158] The resulting strain was named BLT057.

Example 5

Introducing ETF α and ETF β in BLT057

[0159] The ETF genes and the Acdh genes as listed in Table 4 were codon pair optimized for *S. cerevisiae* as described in WO2008/000632 and expressed from yeast promoters and terminators as listed in Table 4.

TABLE 4

Promoters and terminators used for expression of codon pair optimized ETF genes and Acdh genes in <i>S. cerevisiae</i>		
	Promotor	Terminator
Etfa (CpO)	tef1	tdh2
Etfb (CpO)	tdh2	tef1
Acdh64 (AdhE <i>S. aureus</i>)	tdh3	adh
Acdh65 (<i>Clostridium</i>)	tdh3	adh
Acdh66 (EutE <i>E. coli</i>)	tdh3	adh
Acdh67 (lin129 Ec)	tdh3	Adh

[0160] The integration vectors expressing ETF α and ETF β only (pBOL113, [SEQ ID NO:43]) or ETF α and ETF β combined with Acdh64 (pBOL115, [SEQ ID NO:44]), Acdh65 (pBOL116, [SEQ ID NO:45]), Acdh66 (pBOL118, [SEQ ID NO:46]) or Acdh67 (pBOL120, [SEQ ID NO:47]) were synthesized by Geneart AG.

[0161] The vectors, pBOL113, pBOL115, pBOL116, pBOL118 and pBOL120, were linearized with StuI and integrated in the ura3-52 locus of strain BLT057.

[0162] The transformants were grown in YNB (Difco) w/o amino acids+2% galactose to select for uracil prototrophic strains. The strains derived from strain BLT057 with pBOL113/115/116/118/120 integrated in the genome were designated strains: BLT071, BLT072, BLT073, BLT074 and BLT075, respectively.

Example 6

Improved Butanol Production by Expressing Positive Acdh Genes

[0163] Strains BLT071 through BLT075 as prepared in Example 5 were grown in Verduyn medium (Verduyn et al. (1992) Yeast 8: 501-517) in which the ammonium sulphate is replaced by 2 g/l ureum and which further contains 4 wt. % galactose. Cells were grown in 100 ml shake flasks containing 50 ml of medium for 72 hours at 30° C. at 180 rpm in a rotary shaker.

[0164] The butanol concentration was determined in the supernatant of the culture. Samples were analysed on a HS-GC equipped with a flame ionisation detector and an automatic injection system. Column J&W DB-1 length 30 m, id 0.53 mm, df 5 μ m. The following conditions were used: helium as carrier gas with a flow rate of 5 ml/min. Column temperature was set at 110° C. The injector was set at 140° C. and the detector performed at 300° C. The data was obtained using Chromeleon software. Samples were heated at 60° C. for 20 min in the headspace sampler. One (1) ml of the headspace volatiles were automatically injected on the column.

[0165] 1-Butanol production of the various strains was as follows:

[0166] BLT057: 120 mg/l

[0167] BLT071: 450 mg/l

[0168] BLT072: 500 mg/l

[0169] BLT073: 600 mg/l

[0170] BLT074: 670 mg/l

[0171] BLT075: 700 mg/l

[0172] The results show that introduction of electron transfer flavoproteins (ETF alpha and ETF beta) and/or introduction of acetylating acetaldehyde dehydrogenases as identified by a complementation assay of Example 3, increase the butanol production level.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 52

<210> SEQ ID NO 1

<211> LENGTH: 64

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer 5'acs2

-continued

<400> SEQUENCE: 1

atacacaaac agaatacagg aaagtaaadc aatacaataa taaaacagct gaagcttcgt 60

acgc 64

<210> SEQ ID NO 2

<211> LENGTH: 67

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer 3'acs2

<400> SEQUENCE: 2

tttcattacg aaatTTTTct catttaagtt atttcttttt ttgaggcata ggccactagt 60

ggatctg 67

<210> SEQ ID NO 3

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: probe 5'acs2

<400> SEQUENCE: 3

gatattcggt agccgattcc 20

<210> SEQ ID NO 4

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: probe 3'acs2

<400> SEQUENCE: 4

ccgtaacctt ctcgtaatgc 20

<210> SEQ ID NO 5

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: probe ACS2internal

<400> SEQUENCE: 5

cggattcgtc atcagettca 20

<210> SEQ ID NO 6

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: probe KanA

<400> SEQUENCE: 6

cgcacgtcaa gactgtcaag 20

<210> SEQ ID NO 7

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: probe KanB

-continued

<400> SEQUENCE: 7

tcgtatgtga atgctgggtcg

20

<210> SEQ ID NO 8

<211> LENGTH: 64

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer MCS-5'Tadh

<400> SEQUENCE: 8

aaggtagctc tagactagtc ccgggctgca gtcgactcga gcgaatttct tatgatttat

60

gatt

64

<210> SEQ ID NO 9

<211> LENGTH: 39

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer Tadh1-Hind

<400> SEQUENCE: 9

aggaagctta ggctgtgtg gaagaacgat tacaacagg

39

<210> SEQ ID NO 10

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer P1

<400> SEQUENCE: 10

gaattgaagg atatctacat caag

24

<210> SEQ ID NO 11

<211> LENGTH: 25

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer P2

<400> SEQUENCE: 11

cccatctacg gaacctgat caagc

25

<210> SEQ ID NO 12

<211> LENGTH: 26

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: primer P3

<400> SEQUENCE: 12

gatgggtgca ccattaccag gtctag

26

<210> SEQ ID NO 13

<211> LENGTH: 56

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Primer P4

<400> SEQUENCE: 13

-continued

gttctctggt caagttgaag tccattttga ttgatttgac tgtgttattt tgcgtg 56

<210> SEQ ID NO 14
 <211> LENGTH: 25
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer P5

<400> SEQUENCE: 14

gaacaataga gcgacatga ccttg 25

<210> SEQ ID NO 15
 <211> LENGTH: 25
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer P6

<400> SEQUENCE: 15

gacatcagcg tcaccagcct tgatg 25

<210> SEQ ID NO 16
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer P7

<400> SEQUENCE: 16

gattgaaggt ttcaagaaca ggtgatg 27

<210> SEQ ID NO 17
 <211> LENGTH: 25
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer P8

<400> SEQUENCE: 17

ggcgatcaga gttgaaaaaa aaatg 25

<210> SEQ ID NO 18
 <211> LENGTH: 1404
 <212> TYPE: DNA
 <213> ORGANISM: Escherichia coli
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)..(1404)

<400> SEQUENCE: 18

atg aat caa cag gat att gaa cag gtg gtg aaa gcg gta ctg ctg aaa	48
Met Asn Gln Gln Asp Ile Glu Gln Val Val Lys Ala Val Leu Leu Lys	
1 5 10 15	

atg caa agc agt gac acg ccg tcc gcc gcc gtt cat gag atg ggc gtt	96
Met Gln Ser Ser Asp Thr Pro Ser Ala Ala Val His Glu Met Gly Val	
20 25 30	

ttc gcg tcc ctg gat gac gcc gtt gcg gca gcc aaa gtc gcc cag caa	144
Phe Ala Ser Ser Leu Asp Asp Ala Val Ala Ala Ala Lys Val Ala Gln Gln	
35 40 45	

ggg tta aaa agc gtg gca atg cgc cag tta gcc att gct gcc att cgt	192
Gly Leu Lys Ser Val Ala Met Arg Gln Leu Ala Ile Ala Ala Ile Arg	

-continued

50	55	60	
gaa gca ggc gaa aaa cac gcc aga gat tta gcg gaa ctt gcc gtc agt Glu Ala Gly Glu Lys His Ala Arg Asp Leu Ala Glu Leu Ala Val Ser 65 70 75 80			240
gaa acc ggc atg ggg cgc gtt gaa gat aaa ttt gca aaa aac gtc gct Glu Thr Gly Met Gly Arg Val Glu Asp Lys Phe Ala Lys Asn Val Ala 85 90 95			288
cag gcg cgc ggc aca cca ggc gtt gag tgc ctc tct ccg caa gtg ctg Gln Ala Arg Gly Thr Pro Gly Val Glu Cys Leu Ser Pro Gln Val Leu 100 105 110			336
act ggc gac aac ggc ctg acc cta att gaa aac gca ccc tgg ggc gtg Thr Gly Asp Asn Gly Leu Thr Leu Ile Glu Asn Ala Pro Trp Gly Val 115 120 125			384
gtg gct tcg gtg acg cct tcc act aac ccg gcg gca acc gta att aac Val Ala Ser Val Thr Pro Ser Thr Asn Pro Ala Ala Thr Val Ile Asn 130 135 140			432
aac gcc atc agc ctg att gcc gcg ggc aac agc gtc att ttt gcc ccg Asn Ala Ile Ser Leu Ile Ala Ala Gly Asn Ser Val Ile Phe Ala Pro 145 150 155 160			480
cat ccg gcg gcg aaa aaa gtc tcc cag ccg gcg att acg ctg ctc aac His Pro Ala Ala Lys Lys Val Ser Gln Arg Ala Ile Thr Leu Leu Asn 165 170 175			528
cag gcg att gtt gcc gca ggt ggg ccg gaa aac tta ctg gtt act gtg Gln Ala Ile Val Ala Ala Gly Gly Pro Glu Asn Leu Leu Val Thr Val 180 185 190			576
gca aat ccg gat atc gaa acc gcg caa cgc ttg ttc aag ttt ccg ggt Ala Asn Pro Asp Ile Glu Thr Ala Gln Arg Leu Phe Lys Phe Pro Gly 195 200 205			624
atc ggc ctg ctg gtg gta acc ggc ggc gaa gcg gta gta gaa gcg gcg Ile Gly Leu Leu Val Val Thr Gly Gly Glu Ala Val Val Glu Ala Ala 210 215 220			672
cgt aaa cac acc aat aaa cgt ctg att gcc gca ggc gct ggc aac ccg Arg Lys His Thr Asn Lys Arg Leu Ile Ala Ala Gly Ala Gly Asn Pro 225 230 235 240			720
ccg gta gtg gtg gat gaa acc gcc gac ctc gcc cgt gcc gct cag tcc Pro Val Val Val Asp Glu Thr Ala Asp Leu Ala Arg Ala Ala Gln Ser 245 250 255			768
atc gtc aaa ggc gct tct ttc gat aac aac atc att tgt gcc gac gaa Ile Val Lys Gly Ala Ser Phe Asp Asn Asn Ile Ile Cys Ala Asp Glu 260 265 270			816
aag gta ctg att gtt gtt gat agc gta gcc gat gaa ctg atg cgt ctg Lys Val Leu Ile Val Val Asp Ser Val Ala Asp Glu Leu Met Arg Leu 275 280 285			864
atg gaa ggc cag cac gcg gtg aaa ctg acc gca gaa cag gcg cag cag Met Glu Gly Gln His Ala Val Lys Leu Thr Ala Glu Gln Ala Gln Gln 290 295 300			912
ctg caa ccg gtg ttg ctg aaa aat atc gac gag cgc gga aaa ggc acc Leu Gln Pro Val Leu Leu Lys Asn Ile Asp Glu Arg Gly Lys Gly Thr 305 310 315 320			960
gtc agc cgt gac tgg gtt ggt cgc gac gca ggc aaa atc gcg gcg gca Val Ser Arg Asp Trp Val Gly Arg Asp Ala Gly Lys Ile Ala Ala Ala 325 330 335			1008
atc ggc ctt aaa gtt ccg caa gaa acg cgc ctg ctg ttt gtg gaa acc Ile Gly Leu Lys Val Pro Gln Glu Thr Arg Leu Leu Phe Val Glu Thr 340 345 350			1056
acc gca gaa cat ccg ttt gcc gtg act gaa ctg atg atg ccg gtg ttg Thr Ala Glu His Pro Phe Ala Val Thr Glu Leu Met Met Pro Val Leu			1104

-continued

355	360	365	
ccc gtc gtg cgc gtc gcc aac gtg gcg gat gcc att gcg cta gcg gtg			1152
Pro Val Val Arg Val Ala Asn Val Ala Asp Ala Ile Ala Leu Ala Val			
370	375	380	
aaa ctg gaa ggc ggt tgc cac cac acg gcg gca atg cac tcg cgc aac			1200
Lys Leu Glu Gly Gly Cys His His Thr Ala Ala Met His Ser Arg Asn			
385	390	395	400
atc gaa aac atg aac cag atg gcg aat gct att gat acc agc att ttc			1248
Ile Glu Asn Met Asn Gln Met Ala Asn Ala Ile Asp Thr Ser Ile Phe			
405	410	415	
gtt aag aac gga cgc tgc att gcc ggg ctg ggg ctg ggc ggg gaa ggc			1296
Val Lys Asn Gly Pro Cys Ile Ala Gly Leu Gly Leu Gly Gly Glu Gly			
420	425	430	
tgg acc acc atg acc atc acc acg cca acc ggt gaa ggg gta acc agc			1344
Trp Thr Thr Met Thr Ile Thr Thr Pro Thr Gly Glu Gly Val Thr Ser			
435	440	445	
gcg cgt acg ttt gtc cgt ctg cgt cgc tgt gta tta gtc gat gcg ttt			1392
Ala Arg Thr Phe Val Arg Leu Arg Arg Cys Val Leu Val Asp Ala Phe			
450	455	460	
cgc att gtt taa			1404
Arg Ile Val			
465			

<210> SEQ ID NO 19

<211> LENGTH: 467

<212> TYPE: PRT

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 19

Met Asn Gln Gln Asp Ile Glu Gln Val Val Lys Ala Val Leu Leu Lys	
1	15
Met Gln Ser Ser Asp Thr Pro Ser Ala Ala Val His Glu Met Gly Val	
20	30
Phe Ala Ser Leu Asp Asp Ala Val Ala Ala Ala Lys Val Ala Gln Gln	
35	45
Gly Leu Lys Ser Val Ala Met Arg Gln Leu Ala Ile Ala Ala Ile Arg	
50	60
Glu Ala Gly Glu Lys His Ala Arg Asp Leu Ala Glu Leu Ala Val Ser	
65	80
Glu Thr Gly Met Gly Arg Val Glu Asp Lys Phe Ala Lys Asn Val Ala	
85	95
Gln Ala Arg Gly Thr Pro Gly Val Glu Cys Leu Ser Pro Gln Val Leu	
100	110
Thr Gly Asp Asn Gly Leu Thr Leu Ile Glu Asn Ala Pro Trp Gly Val	
115	125
Val Ala Ser Val Thr Pro Ser Thr Asn Pro Ala Ala Thr Val Ile Asn	
130	140
Asn Ala Ile Ser Leu Ile Ala Ala Gly Asn Ser Val Ile Phe Ala Pro	
145	160
His Pro Ala Ala Lys Lys Val Ser Gln Arg Ala Ile Thr Leu Leu Asn	
165	175
Gln Ala Ile Val Ala Ala Gly Gly Pro Glu Asn Leu Leu Val Thr Val	
180	190
Ala Asn Pro Asp Ile Glu Thr Ala Gln Arg Leu Phe Lys Phe Pro Gly	
195	205

-continued

Ile Gly Leu Leu Val Val Thr Gly Gly Glu Ala Val Val Glu Ala Ala
 210 215 220
 Arg Lys His Thr Asn Lys Arg Leu Ile Ala Ala Gly Ala Gly Asn Pro
 225 230 235 240
 Pro Val Val Val Asp Glu Thr Ala Asp Leu Ala Arg Ala Ala Gln Ser
 245 250 255
 Ile Val Lys Gly Ala Ser Phe Asp Asn Asn Ile Ile Cys Ala Asp Glu
 260 265 270
 Lys Val Leu Ile Val Val Asp Ser Val Ala Asp Glu Leu Met Arg Leu
 275 280 285
 Met Glu Gly Gln His Ala Val Lys Leu Thr Ala Glu Gln Ala Gln Gln
 290 295 300
 Leu Gln Pro Val Leu Leu Lys Asn Ile Asp Glu Arg Gly Lys Gly Thr
 305 310 315 320
 Val Ser Arg Asp Trp Val Gly Arg Asp Ala Gly Lys Ile Ala Ala Ala
 325 330 335
 Ile Gly Leu Lys Val Pro Gln Glu Thr Arg Leu Leu Phe Val Glu Thr
 340 345 350
 Thr Ala Glu His Pro Phe Ala Val Thr Glu Leu Met Met Pro Val Leu
 355 360 365
 Pro Val Val Arg Val Ala Asn Val Ala Asp Ala Ile Ala Leu Ala Val
 370 375 380
 Lys Leu Glu Gly Gly Cys His His Thr Ala Ala Met His Ser Arg Asn
 385 390 395 400
 Ile Glu Asn Met Asn Gln Met Ala Asn Ala Ile Asp Thr Ser Ile Phe
 405 410 415
 Val Lys Asn Gly Pro Cys Ile Ala Gly Leu Gly Leu Gly Gly Glu Gly
 420 425 430
 Trp Thr Thr Met Thr Ile Thr Thr Thr Gly Glu Gly Val Thr Ser
 435 440 445
 Ala Arg Thr Phe Val Arg Leu Arg Arg Cys Val Leu Val Asp Ala Phe
 450 455 460
 Arg Ile Val
 465

<210> SEQ ID NO 20

<211> LENGTH: 1401

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: optimised sequence

<400> SEQUENCE: 20

```

atgaaccaac aagatatcga acaagttgtc aaggtgtct tgttgaaaat gcaatcttct    60
gacactccat ctgctgctgt ccacgaaatg ggtgttttcg cttctttgga cgacgctgtt    120
gctgctgccca aggttgctca acaaggtttg aaatctgttg ccatgagaca attggccatt    180
gctgccatca gagaagctgg tgaaaagcat gccagagact tggtgaatt ggctgtctcc    240
gaaaccggta tgggtagagt tgaagacaaa ttcgctaaga acgttgctca agctagaggt    300
actccaggtg tcgaatgttt gtctccacaa gtcttgaccg gtgataatgg tttgactttg    360
attgaaaatg ctccatgggg tgttggtgct tccgtcacc catctaccaa cccagctgct    420

```

-continued

```

actgtcatca acaacgccat ctctttgatt gctgctggta actccgttat cttecgctcca 480
caccagctg ccaagaaggt ttctcaaaga gccatcactc tattgaacca agccattggt 540
gctgctgggt gtccagaaaa ctgtttggtc actgttgcca acccagatat cgaaactgct 600
caaagattat tcaagttccc aggtatcggt ctattagtcg tcaactggtgg tgaagctggt 660
gttgaagctg ccagaaagca caccaacaag agattgattg ctgctggtgc tggtaaccct 720
cctgttggtg tcgatgaaac cgctgatttg gccagagctg ctcaatccat tgtcaagggt 780
gcttcttttc acaacaacat catctgtgct gacgaaaagg ttttgattgt tgttgactcc 840
gttgctgacg aattgatgag attgatggaa ggtcaacatg ccgtaagtt gactgctgaa 900
caagctcaac aattgcaacc agttttgttg aagaacatcg atgaaagagg taagggtacc 960
gtctccagag actgggttgg tagagatgct ggtaagattg ctgctgccat cggtttgaag 1020
gttcacaag aaaccagatt attattcgtc gaaaccaccg ctgaacaccc atttgcgtgc 1080
actgaattga tgatgccagt cttaccagtt gtccgtgttg ctaacgttgc tgacgctatt 1140
gctttggctg tcaaattgga aggtggttgt caccacactg ctgccatgca ctccagaaac 1200
atcgaaaaca tgaaccaaat ggctaacgcc attgacactt ccatctttgt caagaacggt 1260
ccatgtatcg ctgggttggg ttgggtgggt gaagggttga ccaccatgac catcaccacc 1320
ccaactgggt aaggtgtcac ttctgccaga actttcgtca gattacgtcg ttgtgttttg 1380
gtcgatgctt tcagaattgt t 1401

```

```

<210> SEQ ID NO 21
<211> LENGTH: 1410
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(1410)

```

```

<400> SEQUENCE: 21

```

```

atg gaa tca tta gaa ctc gaa caa ctg gta aaa aaa gtt ctc tta gaa 48
Met Glu Ser Leu Glu Leu Glu Gln Leu Val Lys Lys Val Leu Leu Glu
1 5 10 15

aaa tta gca gaa caa aaa gaa gta cca aca aaa aca act aca caa ggc 96
Lys Leu Ala Glu Gln Lys Glu Val Pro Thr Lys Thr Thr Thr Gln Gly
20 25 30

gcg aaa agt ggc gtt ttt gat aca gtt gac gag gct gtt caa gca gca 144
Ala Lys Ser Gly Val Phe Asp Thr Val Asp Glu Ala Val Gln Ala Ala
35 40 45

gtt ata gcg cag aat tgc tat aaa gaa aaa tca ctt gaa gaa cgc cgc 192
Val Ile Ala Gln Asn Cys Tyr Lys Glu Lys Ser Leu Glu Glu Arg Arg
50 55 60

aat gtt gta aaa gca att cgt gaa gca ctt tat cca gaa att gaa aca 240
Asn Val Val Lys Ala Ile Arg Glu Ala Leu Tyr Pro Glu Ile Glu Thr
65 70 75 80

att gcg aca aga gca gtt gca gag act ggt atg gga aat gtg aca gat 288
Ile Ala Thr Arg Ala Val Ala Glu Thr Gly Met Gly Asn Val Thr Asp
85 90 95

aaa att ttg aaa aac acg tta gca atc gaa aaa acg cca ggg gta gaa 336
Lys Ile Leu Lys Asn Thr Leu Ala Ile Glu Lys Thr Pro Gly Val Glu
100 105 110

gat tta tat aca gaa gta gct aca ggt gat aac ggt atg aca cta tat 384
Asp Leu Tyr Thr Glu Val Ala Thr Gly Asp Asn Gly Met Thr Leu Tyr
115 120 125

```

-continued

gaa ctc tct ccg tat ggc gta att ggt gca gta gcg ccg agc aca aac Glu Leu Ser Pro Tyr Gly Val Ile Gly Ala Val Ala Pro Ser Thr Asn 130 135 140	432
cca acg gaa aca ttg att tgt aat tca atc ggt atg ctc gca gct gga Pro Thr Glu Thr Leu Ile Cys Asn Ser Ile Gly Met Leu Ala Ala Gly 145 150 155 160	480
aat gcc gtt ttt tat agc cct cat cca ggg gca aaa aac att tca ctg Asn Ala Val Phe Tyr Ser Pro His Pro Gly Ala Lys Asn Ile Ser Leu 165 170 175	528
tgg ttg att gaa aaa cta aac aca att gtt cgc gat agt tgt ggt ata Trp Leu Ile Glu Lys Leu Asn Thr Ile Val Arg Asp Ser Cys Gly Ile 180 185 190	576
gat aat cta att gtc acc gtg gct aaa cca tcc atc caa gca gct caa Asp Asn Leu Ile Val Thr Val Ala Lys Pro Ser Ile Gln Ala Ala Gln 195 200 205	624
gaa atg atg aac cat cca aaa gta ccg cta ctt gtt att aca ggt ggt Glu Met Met Asn His Pro Lys Val Pro Leu Leu Val Ile Thr Gly Gly 210 215 220	672
ccg ggc gtt gtt ctc caa gcg atg caa tca ggt aaa aaa gtg att gga Pro Gly Val Val Leu Gln Ala Met Gln Ser Gly Lys Lys Val Ile Gly 225 230 235 240	720
gca gga gca ggg aac ccg cct tct att gtt gac gaa aca gct aat atc Ala Gly Ala Gly Asn Pro Pro Ser Ile Val Asp Glu Thr Ala Asn Ile 245 250 255	768
gaa aaa gcg gct gct gac atc gta gac gga gca tct ttt gac cat aat Glu Lys Ala Ala Ala Asp Ile Val Asp Gly Ala Ser Phe Asp His Asn 260 265 270	816
att tta tgt att gct gaa aaa agt gtg gta gct gtt gat agc att gct Ile Leu Cys Ile Ala Glu Lys Ser Val Val Ala Val Asp Ser Ile Ala 275 280 285	864
gat ttc ttg tta ttc caa atg gaa aaa aat ggt gcc ctt cat gtt act Asp Phe Leu Leu Phe Gln Met Glu Lys Asn Gly Ala Leu His Val Thr 290 295 300	912
aat cca agt gat att caa aaa tta gaa aaa gta gcc gtt acc gat aaa Asn Pro Ser Asp Ile Gln Lys Leu Glu Lys Val Ala Val Thr Asp Lys 305 310 315 320	960
ggt gta act aat aaa aaa tta gtc gga aaa agt gca act gaa atc tta Gly Val Thr Asn Lys Lys Leu Val Gly Lys Ser Ala Thr Glu Ile Leu 325 330 335	1008
aaa gaa gca gga ata gct tgt gat ttt aca cca cgt tta atc att gtg Lys Glu Ala Gly Ile Ala Cys Asp Phe Thr Pro Arg Leu Ile Ile Val 340 345 350	1056
gaa acg gag aaa tct cat cca ttt gca aca gta gag cta tta atg cca Glu Thr Glu Lys Ser His Pro Phe Ala Thr Val Glu Leu Leu Met Pro 355 360 365	1104
atc gtt cca gtt gta agg gtg cct gat ttt gac gaa gcc ctt gaa gtg Ile Val Pro Val Val Arg Val Pro Asp Phe Asp Glu Ala Leu Glu Val 370 375 380	1152
gct att gaa ctc gaa caa ggc tta cat cat aca gca aca atg cat tca Ala Ile Glu Leu Glu Gln Gly Leu His His Thr Ala Thr Met His Ser 385 390 395 400	1200
caa aat atc tcg aga tta aac aaa gct gca aga gat atg caa act tcc Gln Asn Ile Ser Arg Leu Asn Lys Ala Ala Arg Asp Met Gln Thr Ser 405 410 415	1248
atc ttt gtc aaa aat ggt ccg tcc ttt gcg gga tta ggc ttt aga gga Ile Phe Val Lys Asn Gly Pro Ser Phe Ala Gly Leu Gly Phe Arg Gly 420 425 430	1296

-continued

```

gaa ggt agt act act ttc act att gca acg cct act gga gaa gga aca 1344
Glu Gly Ser Thr Thr Phe Thr Ile Ala Thr Pro Thr Gly Glu Gly Thr
      435                440                445

```

```

act aca gca cgt cat ttt gct aga cgc cgc cgc tgt gtt tta aca gat 1392
Thr Thr Ala Arg His Phe Ala Arg Arg Arg Arg Cys Val Leu Thr Asp
      450                455                460

```

```

ggt ttt tcg att cgt taa 1410
Gly Phe Ser Ile Arg
465

```

<210> SEQ ID NO 22

<211> LENGTH: 469

<212> TYPE: PRT

<213> ORGANISM: *Listeria innocua*

<400> SEQUENCE: 22

```

Met Glu Ser Leu Glu Leu Glu Gln Leu Val Lys Lys Val Leu Leu Glu
1      5      10      15

Lys Leu Ala Glu Gln Lys Glu Val Pro Thr Lys Thr Thr Thr Gln Gly
      20      25      30

Ala Lys Ser Gly Val Phe Asp Thr Val Asp Glu Ala Val Gln Ala Ala
      35      40      45

Val Ile Ala Gln Asn Cys Tyr Lys Glu Lys Ser Leu Glu Glu Arg Arg
      50      55      60

Asn Val Val Lys Ala Ile Arg Glu Ala Leu Tyr Pro Glu Ile Glu Thr
      65      70      75      80

Ile Ala Thr Arg Ala Val Ala Glu Thr Gly Met Gly Asn Val Thr Asp
      85      90      95

Lys Ile Leu Lys Asn Thr Leu Ala Ile Glu Lys Thr Pro Gly Val Glu
      100     105     110

Asp Leu Tyr Thr Glu Val Ala Thr Gly Asp Asn Gly Met Thr Leu Tyr
      115     120     125

Glu Leu Ser Pro Tyr Gly Val Ile Gly Ala Val Ala Pro Ser Thr Asn
      130     135     140

Pro Thr Glu Thr Leu Ile Cys Asn Ser Ile Gly Met Leu Ala Ala Gly
      145     150     155     160

Asn Ala Val Phe Tyr Ser Pro His Pro Gly Ala Lys Asn Ile Ser Leu
      165     170     175

Trp Leu Ile Glu Lys Leu Asn Thr Ile Val Arg Asp Ser Cys Gly Ile
      180     185     190

Asp Asn Leu Ile Val Thr Val Ala Lys Pro Ser Ile Gln Ala Ala Gln
      195     200     205

Glu Met Met Asn His Pro Lys Val Pro Leu Leu Val Ile Thr Gly Gly
      210     215     220

Pro Gly Val Val Leu Gln Ala Met Gln Ser Gly Lys Lys Val Ile Gly
      225     230     235     240

Ala Gly Ala Gly Asn Pro Pro Ser Ile Val Asp Glu Thr Ala Asn Ile
      245     250     255

Glu Lys Ala Ala Ala Asp Ile Val Asp Gly Ala Ser Phe Asp His Asn
      260     265     270

Ile Leu Cys Ile Ala Glu Lys Ser Val Val Ala Val Asp Ser Ile Ala
      275     280     285

Asp Phe Leu Leu Phe Gln Met Glu Lys Asn Gly Ala Leu His Val Thr

```

-continued

290	295	300
Asn Pro Ser Asp Ile Gln Lys Leu Glu Lys Val Ala Val Thr Asp Lys 305 310 315 320		
Gly Val Thr Asn Lys Lys Leu Val Gly Lys Ser Ala Thr Glu Ile Leu 325 330 335		
Lys Glu Ala Gly Ile Ala Cys Asp Phe Thr Pro Arg Leu Ile Ile Val 340 345 350		
Glu Thr Glu Lys Ser His Pro Phe Ala Thr Val Glu Leu Leu Met Pro 355 360 365		
Ile Val Pro Val Val Arg Val Pro Asp Phe Asp Glu Ala Leu Glu Val 370 375 380		
Ala Ile Glu Leu Glu Gln Gly Leu His His Thr Ala Thr Met His Ser 385 390 395 400		
Gln Asn Ile Ser Arg Leu Asn Lys Ala Ala Arg Asp Met Gln Thr Ser 405 410 415		
Ile Phe Val Lys Asn Gly Pro Ser Phe Ala Gly Leu Gly Phe Arg Gly 420 425 430		
Glu Gly Ser Thr Thr Phe Thr Ile Ala Thr Pro Thr Gly Glu Gly Thr 435 440 445		
Thr Thr Ala Arg His Phe Ala Arg Arg Arg Arg Cys Val Leu Thr Asp 450 455 460		
Gly Phe Ser Ile Arg 465		

<210> SEQ ID NO 23
 <211> LENGTH: 1407
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: optimised sequence

<400> SEQUENCE: 23

```

atggaatctt tggaattgga acaattagtc aagaaggttt tgttggaata attggctgaa    60
caaaaggaag ttccaaccaa gaccaccacc caagggtgcca agtcgggtgt ttctgatacc    120
gtcgatgaag ctgtccaagc tgccgtcatt gctcaaaact gttacaagga aaaatctttg    180
gaagaaagaa gaaacgttgt caaggccatc agagaagctt tatacccaga aatcgaaacc    240
attgctacca gagctgttgc tgaaaccggt atgggtaatg tcaccgataa aatcttgaag    300
aacactttag ctatcgaaaa gactccaggt gttgaagact tgtacactga agttgctacc    360
ggtgacaacg gtatgacttt atacgaatta tctccatacg gtgtcatcgg tctgttgtct    420
ccatctacca acccaactga aactttgac tgtaactcca tcggtatggt ggctgctggt    480
aacgccgttt tctactctcc tcaccaggt gccagaaca tctctttatg gttgattgaa    540
aagttgaaca ctatcgctag agattcttgt ggtattgaca acttgattgt caccgttgcc    600
aagccatcta tccaagctgc tcaagaaatg atgaaccacc caaagggtcc attggtggtc    660
atcactgggt gtccagggtg tgtcttgcaa gctatgcaat ctggtaagaa gggtatcggt    720
gctgggtgct gtaaccctcc atccatcgtt gacgaaaccg ctaacattga aaaggctgct    780
gctgacattg tcgacggtgc ttcctttgac cataatatct tgtgtatcgc tgaaaagtct    840
gtgtgtgccc ttgactccat tgctgacttc ttgtgttccc aaatggaaaa gaacggtgct    900
ttgcacgtca ctaaccatc tgatatccaa aaattggaaa aggttgccgt cactgacaag    960
  
```

-continued

```

gggtgtcacca acaagaaatt gggttggaag tctgccactg aaatcttgaa agaagctggt 1020
attgcttggtg atttcacccc aagattgatac attgtcgaaa ctgaaaagtc ccacccattc 1080
gctactgttg aattgttgat gccaatgtt ccagttgtca gagttccaga cttcgatgaa 1140
gctttggaag ttgccattga attggaacaa ggtctacatc acactgctac catgcactct 1200
caaaacatct ccagattgaa caaggctgcc cgtgacatgc aaacctccat ctttgtcaag 1260
aacgggccat ctttcgctgg tttaggttcc agaggtgaag gttccaccac tttcaccatt 1320
gctactccaa ctggtgaagg tactaccact gcccgctact tcgctagaag aagaagatgt 1380
gtcttgactg atggtttctc cattaga 1407

```

```

<210> SEQ ID NO 24
<211> LENGTH: 1476
<212> TYPE: DNA
<213> ORGANISM: Clostridium kluyveri
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(1476)

```

```

<400> SEQUENCE: 24

```

```

atg gag ata atg gat aag gac tta cag tca ata cag gaa gta aga act 48
Met Glu Ile Met Asp Lys Asp Leu Gln Ser Ile Gln Glu Val Arg Thr
1 5 10 15

ctt ata gca aaa gca aag aaa gct caa gca gaa ttt aaa aat ttt tct 96
Leu Ile Ala Lys Ala Lys Lys Ala Gln Ala Glu Phe Lys Asn Phe Ser
20 25 30

caa gaa gct gta aac aag gta ata gaa aaa ata gct aag gct aca gaa 144
Gln Glu Ala Val Asn Lys Val Ile Glu Lys Ile Ala Lys Ala Thr Glu
35 40 45

gtt gaa gct gta aaa ctt gca aaa ttg gca tat gaa gat aca gga tat 192
Val Glu Ala Val Lys Leu Ala Lys Leu Ala Tyr Glu Asp Thr Gly Tyr
50 55 60

gga aaa tgg gaa gat aaa gta ata aag aat aag ttt tca agt ata gta 240
Gly Lys Trp Glu Asp Lys Val Ile Lys Asn Lys Phe Ser Ser Ile Val
65 70 75 80

gtt tat aac tat att aaa gat ttg aaa acg gtt gga att tta aaa gaa 288
Val Tyr Asn Tyr Ile Lys Asp Leu Lys Thr Val Gly Ile Leu Lys Glu
85 90 95

gac aag gaa aag aaa tta ata gat ata gct gtt cca ctt gga gtt ata 336
Asp Lys Glu Lys Lys Leu Ile Asp Ile Ala Val Pro Leu Gly Val Ile
100 105 110

gca gga ctt ata cct tca act aac cca act tca aca gca ata ttc aag 384
Ala Gly Leu Ile Pro Ser Thr Asn Pro Thr Ser Thr Ala Ile Phe Lys
115 120 125

gta tta ata gca tta aag gca gga aat gca ata gta ttc tca cca cat 432
Val Leu Ile Ala Leu Lys Ala Gly Asn Ala Ile Val Phe Ser Pro His
130 135 140

cca aca gca gta aga agt att aca gaa act gta aag ata atg cag aaa 480
Pro Thr Ala Val Arg Ser Ile Thr Glu Thr Val Lys Ile Met Gln Lys
145 150 155 160

gct gca gta gaa gca gga gca cca gat gga tta atc caa tgt atg tca 528
Ala Ala Val Glu Ala Gly Ala Pro Asp Gly Leu Ile Gln Cys Met Ser
165 170 175

ata ttg aca gta gaa ggt act gct gaa ttg atg aag aat aag gat aca 576
Ile Leu Thr Val Glu Gly Thr Ala Glu Leu Met Lys Asn Lys Asp Thr
180 185 190

```

-continued

gca ctt atc ctt gca aca ggt gga gaa gga atg gta aga gca gct tac	624
Ala Leu Ile Leu Ala Thr Gly Gly Glu Gly Met Val Arg Ala Ala Tyr	
195 200 205	
agt tca gga aca cca gct ata gga gtt gga cct gga aac ggc cca tgc	672
Ser Ser Gly Thr Pro Ala Ile Gly Val Gly Pro Gly Asn Gly Pro Cys	
210 215 220	
ttt att gaa aga aca gca gat att cct aca gca gta aga aaa gta ata	720
Phe Ile Glu Arg Thr Ala Asp Ile Pro Thr Ala Val Arg Lys Val Ile	
225 230 235 240	
ggc agt gat act ttt gat aat gga gta ata tgt gct tca gaa caa tca	768
Gly Ser Asp Thr Phe Asp Asn Gly Val Ile Cys Ala Ser Glu Gln Ser	
245 250 255	
ata ata gca gag aca gta aag aaa gca gag ata att gaa gaa ttc aag	816
Ile Ile Ala Glu Thr Val Lys Lys Ala Glu Ile Ile Glu Glu Phe Lys	
260 265 270	
aga caa aaa gga tat ttc tta aat gca gaa gaa tca gaa aaa gta ggc	864
Arg Gln Lys Gly Tyr Phe Leu Asn Ala Glu Glu Ser Glu Lys Val Gly	
275 280 285	
aag att tta tta aga gct aat gga aca cca aac cca gca ata gta gga	912
Lys Ile Leu Leu Arg Ala Asn Gly Thr Pro Asn Pro Ala Ile Val Gly	
290 295 300	
aaa gat gtt caa gca tta gca aaa tta gca gga ata agc ata cca agc	960
Lys Asp Val Gln Ala Leu Ala Lys Leu Ala Gly Ile Ser Ile Pro Ser	
305 310 315 320	
gat gcg gta ata tta ctt tca gag cag aca gat gtg agt cca aag aac	1008
Asp Ala Val Ile Leu Leu Ser Glu Gln Thr Asp Val Ser Pro Lys Asn	
325 330 335	
cct tat gca aag gaa aaa tta gct cca gta ctt gca ttc tat aca gta	1056
Pro Tyr Ala Lys Glu Lys Leu Ala Pro Val Leu Ala Phe Tyr Thr Val	
340 345 350	
gaa gac tgg cat gaa gca tgt gaa aaa tcc tta gca ctt ctt cat aac	1104
Glu Asp Trp His Glu Ala Cys Glu Lys Ser Leu Ala Leu Leu His Asn	
355 360 365	
caa gga agt gga cat aca tta ata att cac tca cag aat gaa gaa atc	1152
Gln Gly Ser Gly His Thr Leu Ile Ile His Ser Gln Asn Glu Glu Ile	
370 375 380	
ata aga gaa ttc gca ttg aag aaa cca gta tca aga ata ctt gta aat	1200
Ile Arg Glu Phe Ala Leu Lys Lys Pro Val Ser Arg Ile Leu Val Asn	
385 390 395 400	
tca cct gga tca ctt gga gga ata ggt gga gct aca aat ctt gta cca	1248
Ser Pro Gly Ser Leu Gly Gly Ile Gly Gly Ala Thr Asn Leu Val Pro	
405 410 415	
tca ctt aca tta ggc tgt gga gca gta ggt gga agt gca act tca gat	1296
Ser Leu Thr Leu Gly Cys Gly Ala Val Gly Gly Ser Ala Thr Ser Asp	
420 425 430	
aac gta gga cca gaa aac tta ttc aac ata aga aaa gta gct tat gga	1344
Asn Val Gly Pro Glu Asn Leu Phe Asn Ile Arg Lys Val Ala Tyr Gly	
435 440 445	
act acg aca gta gaa gaa ata aga gaa gct ttt ggt gta gga gca gct	1392
Thr Thr Thr Val Glu Glu Ile Arg Glu Ala Phe Gly Val Gly Ala Ala	
450 455 460	
tca tca agt gca cca gca gaa cca gaa gat aat gaa gat gta cag gct	1440
Ser Ser Ser Ala Pro Ala Glu Pro Glu Asp Asn Glu Asp Val Gln Ala	
465 470 475 480	
ata gta aaa gct ata atg gct aaa tta aat ctt taa	1476
Ile Val Lys Ala Ile Met Ala Lys Leu Asn Leu	
485 490	

-continued

```

<210> SEQ ID NO 25
<211> LENGTH: 491
<212> TYPE: PRT
<213> ORGANISM: Clostridium kluyveri

<400> SEQUENCE: 25

Met Glu Ile Met Asp Lys Asp Leu Gln Ser Ile Gln Glu Val Arg Thr
1          5          10          15

Leu Ile Ala Lys Ala Lys Lys Ala Gln Ala Glu Phe Lys Asn Phe Ser
20          25          30

Gln Glu Ala Val Asn Lys Val Ile Glu Lys Ile Ala Lys Ala Thr Glu
35          40          45

Val Glu Ala Val Lys Leu Ala Lys Leu Ala Tyr Glu Asp Thr Gly Tyr
50          55          60

Gly Lys Trp Glu Asp Lys Val Ile Lys Asn Lys Phe Ser Ser Ile Val
65          70          75          80

Val Tyr Asn Tyr Ile Lys Asp Leu Lys Thr Val Gly Ile Leu Lys Glu
85          90          95

Asp Lys Glu Lys Lys Leu Ile Asp Ile Ala Val Pro Leu Gly Val Ile
100         105         110

Ala Gly Leu Ile Pro Ser Thr Asn Pro Thr Ser Thr Ala Ile Phe Lys
115         120         125

Val Leu Ile Ala Leu Lys Ala Gly Asn Ala Ile Val Phe Ser Pro His
130         135         140

Pro Thr Ala Val Arg Ser Ile Thr Glu Thr Val Lys Ile Met Gln Lys
145         150         155         160

Ala Ala Val Glu Ala Gly Ala Pro Asp Gly Leu Ile Gln Cys Met Ser
165         170         175

Ile Leu Thr Val Glu Gly Thr Ala Glu Leu Met Lys Asn Lys Asp Thr
180         185         190

Ala Leu Ile Leu Ala Thr Gly Gly Glu Gly Met Val Arg Ala Ala Tyr
195         200         205

Ser Ser Gly Thr Pro Ala Ile Gly Val Gly Pro Gly Asn Gly Pro Cys
210         215         220

Phe Ile Glu Arg Thr Ala Asp Ile Pro Thr Ala Val Arg Lys Val Ile
225         230         235         240

Gly Ser Asp Thr Phe Asp Asn Gly Val Ile Cys Ala Ser Glu Gln Ser
245         250         255

Ile Ile Ala Glu Thr Val Lys Lys Ala Glu Ile Ile Glu Glu Phe Lys
260         265         270

Arg Gln Lys Gly Tyr Phe Leu Asn Ala Glu Glu Ser Glu Lys Val Gly
275         280         285

Lys Ile Leu Leu Arg Ala Asn Gly Thr Pro Asn Pro Ala Ile Val Gly
290         295         300

Lys Asp Val Gln Ala Leu Ala Lys Leu Ala Gly Ile Ser Ile Pro Ser
305         310         315         320

Asp Ala Val Ile Leu Leu Ser Glu Gln Thr Asp Val Ser Pro Lys Asn
325         330         335

Pro Tyr Ala Lys Glu Lys Leu Ala Pro Val Leu Ala Phe Tyr Thr Val
340         345         350

Glu Asp Trp His Glu Ala Cys Glu Lys Ser Leu Ala Leu Leu His Asn
355         360         365

```

-continued

Gln Gly Ser Gly His Thr Leu Ile Ile His Ser Gln Asn Glu Glu Ile
 370 375 380
 Ile Arg Glu Phe Ala Leu Lys Lys Pro Val Ser Arg Ile Leu Val Asn
 385 390 395 400
 Ser Pro Gly Ser Leu Gly Gly Ile Gly Gly Ala Thr Asn Leu Val Pro
 405 410 415
 Ser Leu Thr Leu Gly Cys Gly Ala Val Gly Gly Ser Ala Thr Ser Asp
 420 425 430
 Asn Val Gly Pro Glu Asn Leu Phe Asn Ile Arg Lys Val Ala Tyr Gly
 435 440 445
 Thr Thr Thr Val Glu Glu Ile Arg Glu Ala Phe Gly Val Gly Ala Ala
 450 455 460
 Ser Ser Ser Ala Pro Ala Glu Pro Glu Asp Asn Glu Asp Val Gln Ala
 465 470 475 480
 Ile Val Lys Ala Ile Met Ala Lys Leu Asn Leu
 485 490

<210> SEQ ID NO 26
 <211> LENGTH: 1473
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: optimised sequence

<400> SEQUENCE: 26

```

atggaaatca tggacaagga ttgcaatcc atccaagaag ttagaacttt gattgccaaag    60
gccaaagaag ctcaagctga attcaagaac ttttccaag aagctgttaa caaggtcatc    120
gaaaagatcg ccaaggctac tgaagttgaa gctgtcaaat tggccaaatt ggcttacgaa    180
gacaccggtt acggtaaatg ggaagacaag gtcacaaaga acaaatcttc ctccattggt    240
gtctacaact acatcaagga ttgaagacc gttggtatct tgaaggaaga caaggaaaag    300
aaattgattg acattgctgt cccattaggt gtcattgctg gtttgattcc atctaccaac    360
ccaacttcca ctgccatttt caaggtcttg attgcttga aggctggtaa cgccattgtc    420
ttctctccac acccaactgc tgtccgttcc atcactgaaa ccgttaagat catgcaaaaag    480
gctgctgttg aagctggtgc tccagatggt ttgatccaat gtatgtccat tttgaccgtt    540
gaaggtactg ctgaattgat gaagaacaag gacaccgctt tgatcttggc taccggtggt    600
gaaggtatgg ttagagctgc ttactcctct ggtactccag ccacggtgt cggtccaggt    660
aacggtccat gtttcatcga aagaactgct gacattccaa ctgctgttag aaagggtatc    720
ggttctgaca ctttcgacaa cgggtgtcatc tgtgcttctg aacaatccat cattgctgaa    780
accgtcaaga aggctgaaat catcgaagaa ttcaagagac aaaagggtta cttcttgaat    840
gctgaagaat ctgaaaaggt tggttaagatt ctattacgtg ccaacggtac tccaaacca    900
gccatcgttg gtaaggatgt ccaagctttg gccaaattgg ctggtatttc cattccatct    960
gatgctgtta tcttactatc cgaacaaacc gatgtttctc ctaaaaatcc atacgctaag    1020
gaaaaattgg ctccagtcct ggctttctac accgtcgaag actggcatga agcttgtgaa    1080
aagtcttttg ctttattgca caaccaaggt tctggtcaca ctttgatcat ccactctcaa    1140
aacgaagaaa tcattagaga atttgctttg aagaagcctg tttccagaat tttggttaac    1200
tctccaggtt ctttgggtgg tatcgggtgt gctaccaact tagtcccatc tttgacttta    1260
  
```

-continued

```

ggttgtggtg ctgttggtgg ttctgccacc tctgacaacg ttggtccaga aaacttggtc 1320
aacatcagaa aggttgctta cgggtaccacc accgtcgaag aaatcagaga agcttttcggt 1380
gtcggtgctg cttcttcttc tgctccagct gaaccagaag acaacgaaga tgttcaagcc 1440
attgttaagg ccatcatggc caaattgaac ttg 1473

```

```

<210> SEQ ID NO 27
<211> LENGTH: 2610
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(2610)

```

```

<400> SEQUENCE: 27

```

```

atg tta act ata cct gaa aaa gaa aat cgt gga tcg aaa gaa caa gaa 48
Met Leu Thr Ile Pro Glu Lys Glu Asn Arg Gly Ser Lys Glu Gln Glu
1 5 10 15

gtg gca att atg att gat gct cta gct gac aaa ggg aaa aaa gca tta 96
Val Ala Ile Met Ile Asp Ala Leu Ala Asp Lys Gly Lys Lys Ala Leu
20 25 30

gaa gca tta tct aaa aag tca caa gaa gaa att gat cat att gtt cat 144
Glu Ala Leu Ser Lys Lys Ser Glu Glu Ile Asp His Ile Val His
35 40 45

caa atg agc tta gca gct gtt gat caa cat atg gtg cta gca aaa tta 192
Gln Met Ser Leu Ala Ala Val Asp Gln His Met Val Leu Ala Lys Leu
50 55 60

gca cat gaa gaa act gga aga ggt ata tac gaa gat aaa gcg att aaa 240
Ala His Glu Glu Thr Gly Arg Gly Ile Tyr Glu Asp Lys Ala Ile Lys
65 70 75 80

aat tta tac gct tct gaa tat ata tgg aat tca ata aaa gac aat aag 288
Asn Leu Tyr Ala Ser Glu Tyr Ile Trp Asn Ser Ile Lys Asp Asn Lys
85 90 95

aca gta ggg att att ggt gaa gat aaa gaa aaa gga tta acg tat gta 336
Thr Val Gly Ile Ile Gly Glu Asp Lys Glu Lys Gly Leu Thr Tyr Val
100 105 110

gcg gaa cca att ggt gtt att tgt ggt gtt acg cca aca aca aat cct 384
Ala Glu Pro Ile Gly Val Ile Cys Gly Val Thr Pro Thr Thr Asn Pro
115 120 125

acg tcg aca act att ttt aaa gcg atg att gca att aag aca gga aat 432
Thr Ser Thr Thr Ile Phe Lys Ala Met Ile Ala Ile Lys Thr Gly Asn
130 135 140

cca atc att ttt gca ttc cat cca agt gca caa gaa tcg tcg aag cgt 480
Pro Ile Ile Phe Ala Phe His Pro Ser Ala Gln Glu Ser Ser Lys Arg
145 150 155 160

gca gca gaa gtt gta tta gaa gcg gca atg aag gca ggt gca cct aaa 528
Ala Ala Glu Val Val Leu Glu Ala Ala Met Lys Ala Gly Ala Pro Lys
165 170 175

gat att att cag tgg att gaa gtg cct tct atc gaa gca aca aaa caa 576
Asp Ile Ile Gln Trp Ile Glu Val Pro Ser Ile Glu Ala Thr Lys Gln
180 185 190

tta atg aat cac aaa ggt att gca tta gtt cta gca aca ggt ggt tcg 624
Leu Met Asn His Lys Gly Ile Ala Leu Val Leu Ala Thr Gly Gly Ser
195 200 205

ggc atg gtt aag tct gca tat tca act ggc aaa ccg gca tta ggt gtg 672
Gly Met Val Lys Ser Ala Tyr Ser Thr Gly Lys Pro Ala Leu Gly Val
210 215 220

```

-continued

gga cca ggt aac gtg ccg tct tac att gaa aaa aca gca cac att aaa	720
Gly Pro Gly Asn Val Pro Ser Tyr Ile Glu Lys Thr Ala His Ile Lys	
225 230 235 240	
cgt gca gta aat gat atc att ggt tca aaa aca ttt gat aat ggt atg	768
Arg Ala Val Asn Asp Ile Ile Gly Ser Lys Thr Phe Asp Asn Gly Met	
245 250 255	
att tgt gct tct gaa caa gtt gta gtc att gat aaa gaa att tat aaa	816
Ile Cys Ala Ser Glu Gln Val Val Val Ile Asp Lys Glu Ile Tyr Lys	
260 265 270	
gat gtt act aat gaa ttt aaa gca cat caa gca tac ttt gtt aaa aaa	864
Asp Val Thr Asn Glu Phe Lys Ala His Gln Ala Tyr Phe Val Lys Lys	
275 280 285	
gat gaa tta caa cgc tta gaa aat gca att atg aat gaa caa aaa aca	912
Asp Glu Leu Gln Arg Leu Glu Asn Ala Ile Met Asn Glu Gln Lys Thr	
290 295 300	
ggg att aag cct gat att gtc ggt aaa tct gca gtt gaa ata gct gaa	960
Gly Ile Lys Pro Asp Ile Val Gly Lys Ser Ala Val Glu Ile Ala Glu	
305 310 315 320	
tta gca ggt ata cct gtc ccc gaa aat aca aaa ctt atc ata gcc gaa	1008
Leu Ala Gly Ile Pro Val Pro Glu Asn Thr Lys Leu Ile Ile Ala Glu	
325 330 335	
att agc ggt gta ggt tca gac tat ccg tta tct cgt gaa aaa tta tct	1056
Ile Ser Gly Val Gly Ser Asp Tyr Pro Leu Ser Arg Glu Lys Leu Ser	
340 345 350	
cca gta tta gcc tta gta aaa gcc caa tct aca aaa caa gca ttt caa	1104
Pro Val Leu Ala Leu Val Lys Ala Gln Ser Thr Lys Gln Ala Phe Gln	
355 360 365	
att tgt gaa gac aca cta cat ttt ggt gga tta gga cac aca gcc gtt	1152
Ile Cys Glu Asp Thr Leu His Phe Gly Gly Leu Gly His Thr Ala Val	
370 375 380	
atc cat aca gaa gat gaa aca tta caa aaa gat ttt gga cta aga atg	1200
Ile His Thr Glu Asp Glu Thr Leu Gln Lys Asp Phe Gly Leu Arg Met	
385 390 395 400	
aaa gct tgt cgt gta ctt gta aat aca cca tca gcg gtt gga ggt att	1248
Lys Ala Cys Arg Val Leu Val Asn Thr Pro Ser Ala Val Gly Gly Ile	
405 410 415	
ggg gat atg tat aac gaa ttg att ccg tct tta aca tta ggt tgt ggt	1296
Gly Asp Met Tyr Asn Glu Leu Ile Pro Ser Leu Thr Leu Gly Cys Gly	
420 425 430	
tcg tac ggt aga aac tca att tca cat aat gtt agt gcg aca gat tta	1344
Ser Tyr Gly Arg Asn Ser Ile Ser His Asn Val Ser Ala Thr Asp Leu	
435 440 445	
tta aac att aaa acg att gct aaa cga cgt aat aat act caa att ttc	1392
Leu Asn Ile Lys Thr Ile Ala Lys Arg Arg Asn Asn Thr Gln Ile Phe	
450 455 460	
aag gtg cct gct caa att tat ttt gaa gaa aat gca atc atg agt cta	1440
Lys Val Pro Ala Gln Ile Tyr Phe Glu Glu Asn Ala Ile Met Ser Leu	
465 470 475 480	
aca aca atg gac aag att gaa aaa gtg atg att gtc tgt gac cct ggt	1488
Thr Thr Met Asp Lys Ile Glu Lys Val Met Ile Val Cys Asp Pro Gly	
485 490 495	
atg gta gaa ttc ggt tat aca aaa aca gtt gag aat gta tta aga caa	1536
Met Val Glu Phe Gly Tyr Thr Lys Thr Val Glu Asn Val Leu Arg Gln	
500 505 510	
aga acg gaa cag cct caa att aaa ata ttt agc gaa gtc gaa ccg aac	1584
Arg Thr Glu Gln Pro Gln Ile Lys Ile Phe Ser Glu Val Glu Pro Asn	
515 520 525	

-continued

cca tca act aat aca gta tat aaa ggt ctg gaa atg atg gtt gat ttc	1632
Pro Ser Thr Asn Thr Val Tyr Lys Gly Leu Glu Met Met Val Asp Phe	
530 535 540	
caa cca gat aca atc att gca ctt ggt ggt ggt tca gcg atg gat gct	1680
Gln Pro Asp Thr Ile Ile Ala Leu Gly Gly Gly Ser Ala Met Asp Ala	
545 550 555 560	
gca aaa gca atg tgg atg ttc ttt gaa cac cct gag aca tca ttc ttc	1728
Ala Lys Ala Met Trp Met Phe Phe Glu His Pro Glu Thr Ser Phe Phe	
565 570 575	
ggt gct aaa caa aag ttc cta gac atc ggt aaa cgt act tat aaa ata	1776
Gly Ala Lys Gln Lys Phe Leu Asp Ile Gly Lys Arg Thr Tyr Lys Ile	
580 585 590	
ggc atg cct gaa aat gcg acg ttc att tgt atc cct acg aca tca ggt	1824
Gly Met Pro Glu Asn Ala Thr Phe Ile Cys Ile Pro Thr Thr Ser Gly	
595 600 605	
aca ggt tca gaa gta aca cca ttt gca gtt atc aca gat agt gaa aca	1872
Thr Gly Ser Glu Val Thr Pro Phe Ala Val Ile Thr Asp Ser Glu Thr	
610 615 620	
aat gta aaa tat ccg ttg gct gat ttt gct tta aca cct gac gtt gca	1920
Asn Val Lys Tyr Pro Leu Ala Asp Phe Ala Leu Thr Pro Asp Val Ala	
625 630 635 640	
att att gac cct caa ttt gtg atg agt gtg cca aaa agc gtt aca gca	1968
Ile Ile Asp Pro Gln Phe Val Met Ser Val Pro Lys Ser Val Thr Ala	
645 650 655	
gat aca gga atg gat gta cta acg cat gca atg gaa tca tat gta tct	2016
Asp Thr Gly Met Asp Val Leu Thr His Ala Met Glu Ser Tyr Val Ser	
660 665 670	
gta atg gct tca gac tat aca aga ggt ttg agt cta caa gcg att aaa	2064
Val Met Ala Ser Asp Tyr Thr Arg Gly Leu Ser Leu Gln Ala Ile Lys	
675 680 685	
ttg acg ttc gaa tat tta aaa tca tct gtt gaa aag ggt gat aaa gtt	2112
Leu Thr Phe Glu Tyr Leu Lys Ser Ser Val Glu Lys Gly Asp Lys Val	
690 695 700	
tca aga gag aaa atg cat aac gca tca act ttg gct ggt atg gca ttt	2160
Ser Arg Glu Lys Met His Asn Ala Ser Thr Leu Ala Gly Met Ala Phe	
705 710 715 720	
gca aat gca ttc tta ggc att gca cac tca att gca cat aaa att ggt	2208
Ala Asn Ala Phe Leu Gly Ile Ala His Ser Ile Ala His Lys Ile Gly	
725 730 735	
ggc gaa tat ggt att ccg cat ggt aga gcg aat gcg ata tta cta ccg	2256
Gly Glu Tyr Gly Ile Pro His Gly Arg Ala Asn Ala Ile Leu Leu Pro	
740 745 750	
cat att atc cgt tat aat gcc aaa gac ccg caa aaa cat gca tta ttc	2304
His Ile Ile Arg Tyr Asn Ala Lys Asp Pro Gln Lys His Ala Leu Phe	
755 760 765	
cct aaa tat gag ttc ttc aga gca gat aca gat tat gca gat att gcc	2352
Pro Lys Tyr Glu Phe Phe Arg Ala Asp Thr Asp Tyr Ala Asp Ile Ala	
770 775 780	
aaa ttc tta gga tta aaa ggg aat acg aca gaa gca ctc gta gaa tca	2400
Lys Phe Leu Gly Leu Lys Gly Asn Thr Thr Glu Ala Leu Val Glu Ser	
785 790 795 800	
tta gct aaa gct gtc tac gaa tta ggt caa tca gtc gga att gaa atg	2448
Leu Ala Lys Ala Val Tyr Glu Leu Gly Gln Ser Val Gly Ile Glu Met	
805 810 815	
aat ttg aaa tca caa ggt gtg tct gaa gaa gaa tta aat gaa tca att	2496
Asn Leu Lys Ser Gln Gly Val Ser Glu Glu Glu Leu Asn Glu Ser Ile	
820 825 830	

-continued

gat aga atg gca gag ctc gca ttt gaa gat caa tgt aca act gct aat	2544
Asp Arg Met Ala Glu Leu Ala Phe Glu Asp Gln Cys Thr Thr Ala Asn	
835 840 845	
cct aaa gaa gca cta atc agt gaa atc aaa gat atc att caa aca tca	2592
Pro Lys Glu Ala Leu Ile Ser Glu Ile Lys Asp Ile Ile Gln Thr Ser	
850 855 860	
tat gat tat aag caa taa	2610
Tyr Asp Tyr Lys Gln	
865	

<210> SEQ ID NO 28

<211> LENGTH: 869

<212> TYPE: PRT

<213> ORGANISM: Staphylococcus aureus

<400> SEQUENCE: 28

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

-continued

Gly	Ile	Lys	Pro	Asp	Ile	Val	Gly	Lys	Ser	Ala	Val	Glu	Ile	Ala	Glu	305	310	315	320
Leu	Ala	Gly	Ile	Pro	Val	Pro	Glu	Asn	Thr	Lys	Leu	Ile	Ile	Ala	Glu	325	330	335	
Ile	Ser	Gly	Val	Gly	Ser	Asp	Tyr	Pro	Leu	Ser	Arg	Glu	Lys	Leu	Ser	340	345	350	
Pro	Val	Leu	Ala	Leu	Val	Lys	Ala	Gln	Ser	Thr	Lys	Gln	Ala	Phe	Gln	355	360	365	
Ile	Cys	Glu	Asp	Thr	Leu	His	Phe	Gly	Gly	Leu	Gly	His	Thr	Ala	Val	370	375	380	
Ile	His	Thr	Glu	Asp	Glu	Thr	Leu	Gln	Lys	Asp	Phe	Gly	Leu	Arg	Met	385	390	395	400
Lys	Ala	Cys	Arg	Val	Leu	Val	Asn	Thr	Pro	Ser	Ala	Val	Gly	Gly	Ile	405	410	415	
Gly	Asp	Met	Tyr	Asn	Glu	Leu	Ile	Pro	Ser	Leu	Thr	Leu	Gly	Cys	Gly	420	425	430	
Ser	Tyr	Gly	Arg	Asn	Ser	Ile	Ser	His	Asn	Val	Ser	Ala	Thr	Asp	Leu	435	440	445	
Leu	Asn	Ile	Lys	Thr	Ile	Ala	Lys	Arg	Arg	Asn	Asn	Thr	Gln	Ile	Phe	450	455	460	
Lys	Val	Pro	Ala	Gln	Ile	Tyr	Phe	Glu	Glu	Asn	Ala	Ile	Met	Ser	Leu	465	470	475	480
Thr	Thr	Met	Asp	Lys	Ile	Glu	Lys	Val	Met	Ile	Val	Cys	Asp	Pro	Gly	485	490	495	
Met	Val	Glu	Phe	Gly	Tyr	Thr	Lys	Thr	Val	Glu	Asn	Val	Leu	Arg	Gln	500	505	510	
Arg	Thr	Glu	Gln	Pro	Gln	Ile	Lys	Ile	Phe	Ser	Glu	Val	Glu	Pro	Asn	515	520	525	
Pro	Ser	Thr	Asn	Thr	Val	Tyr	Lys	Gly	Leu	Glu	Met	Met	Val	Asp	Phe	530	535	540	
Gln	Pro	Asp	Thr	Ile	Ile	Ala	Leu	Gly	Gly	Gly	Ser	Ala	Met	Asp	Ala	545	550	555	560
Ala	Lys	Ala	Met	Trp	Met	Phe	Phe	Glu	His	Pro	Glu	Thr	Ser	Phe	Phe	565	570	575	
Gly	Ala	Lys	Gln	Lys	Phe	Leu	Asp	Ile	Gly	Lys	Arg	Thr	Tyr	Lys	Ile	580	585	590	
Gly	Met	Pro	Glu	Asn	Ala	Thr	Phe	Ile	Cys	Ile	Pro	Thr	Thr	Ser	Gly	595	600	605	
Thr	Gly	Ser	Glu	Val	Thr	Pro	Phe	Ala	Val	Ile	Thr	Asp	Ser	Glu	Thr	610	615	620	
Asn	Val	Lys	Tyr	Pro	Leu	Ala	Asp	Phe	Ala	Leu	Thr	Pro	Asp	Val	Ala	625	630	635	640
Ile	Ile	Asp	Pro	Gln	Phe	Val	Met	Ser	Val	Pro	Lys	Ser	Val	Thr	Ala	645	650	655	
Asp	Thr	Gly	Met	Asp	Val	Leu	Thr	His	Ala	Met	Glu	Ser	Tyr	Val	Ser	660	665	670	
Val	Met	Ala	Ser	Asp	Tyr	Thr	Arg	Gly	Leu	Ser	Leu	Gln	Ala	Ile	Lys	675	680	685	
Leu	Thr	Phe	Glu	Tyr	Leu	Lys	Ser	Ser	Val	Glu	Lys	Gly	Asp	Lys	Val	690	695	700	

-continued

Ser Arg Glu Lys Met His Asn Ala Ser Thr Leu Ala Gly Met Ala Phe
 705 710 715 720
 Ala Asn Ala Phe Leu Gly Ile Ala His Ser Ile Ala His Lys Ile Gly
 725 730 735
 Gly Glu Tyr Gly Ile Pro His Gly Arg Ala Asn Ala Ile Leu Leu Pro
 740 745 750
 His Ile Ile Arg Tyr Asn Ala Lys Asp Pro Gln Lys His Ala Leu Phe
 755 760 765
 Pro Lys Tyr Glu Phe Phe Arg Ala Asp Thr Asp Tyr Ala Asp Ile Ala
 770 775 780
 Lys Phe Leu Gly Leu Lys Gly Asn Thr Thr Glu Ala Leu Val Glu Ser
 785 790 795 800
 Leu Ala Lys Ala Val Tyr Glu Leu Gly Gln Ser Val Gly Ile Glu Met
 805 810 815
 Asn Leu Lys Ser Gln Gly Val Ser Glu Glu Glu Leu Asn Glu Ser Ile
 820 825 830
 Asp Arg Met Ala Glu Leu Ala Phe Glu Asp Gln Cys Thr Thr Ala Asn
 835 840 845
 Pro Lys Glu Ala Leu Ile Ser Glu Ile Lys Asp Ile Ile Gln Thr Ser
 850 855 860
 Tyr Asp Tyr Lys Gln
 865

<210> SEQ ID NO 29

<211> LENGTH: 2607

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: optimised sequence

<400> SEQUENCE: 29

```

atgttgacca ttccagaaaa ggaaaacaga ggttccaagg aacaagaagt tgccatcatg      60
attgatgctt tagctgacaa aggtaagaag gctttggaag ctttgtccaa gaagtctcaa      120
gaagaaattg accacattgt ccaccaaattg tccttggtctg ctgttgacca acacatgggt      180
ttggccaagt tggctcatga agaaaccggt agaggatatc acgaagacaa ggctatcaag      240
aacttatatc cctctgaata catctggaac tccatcaagg acaacaagac tgttggtatc      300
attggtgaag acaaagaaaa gggtttgacc tacgttgctg aaccaattgg tgtcatctgt      360
ggtgtcactc caaccaccaa cccaacttct accaccatct tcaaggctat gattgccatc      420
aagactggta acccaattat ttctgctttc caccatctg ctcaagaatc ttccaagaga      480
gctgctgaag ttgttttga agctgccatg aaggctgggt ctccaaagga tatcatccaa      540
tggattgaag ttccatccat tgaagctacc aagcaattga tgaaccacaa gggatttgct      600
ttagtcttgg ctaccggtgg ttctggtatg gttaagtctg cttactccac tggtaaacca      660
gctttgggtg ttggtccagg taacgttcca tottacatcg aaaagactgc tcatatcaag      720
cgtgctgtca acgatatcat cggttccaag actttcgata atggtatgat ctgtgcttct      780
gaacaagttg ttgtcattga caaggaaatc tacaaggatg tcaccaatga attcaaggct      840
caccaagcct acttcgtcaa gaaggacgaa ttacaaagat tagaaaacgc catcatgaac      900
gaacaaaaga ctggtatcaa gccagatata gttggtaagt ctgctgttga aattgctgaa      960
ttggccggta tcccagttcc agaaaacacc aaattgatca ttgctgaaat ctccggtgtc     1020

```

-continued

```

ggttctgact acccattgtc cagagaaaag ttgtctccag ttttggcttt agtcaaggct 1080
caatctacca agcaagcttt ccaaattctgt gaagacactt tgcacttcgg tggtttaggt 1140
cacactgctg ttatccacac tgaagacgaa actttgcaa aggatttcgg tctaagaatg 1200
aaggcttgct gtgttttggc caacactcca tctgctgttg gtggtatcgg tgacatgtac 1260
aacgaattga ttccatcctt gactttgggt tgtggttctt acggtagaaa ctccatctcc 1320
cacaacgtct ctgctaccga tttgttgaac atcaagacca tggccaagag aagaacaac 1380
actcaaatct tcaaggttcc agctcaaatc tatttcgaag aaaacgctat catgtccttg 1440
accaccatgg acaagattga aaaggtcatg atcgtttgg acccaggtat ggttgaattt 1500
ggttacacca aaaccgtcga aaacgtctta cgtcaaagaa ctgaacaacc tcaaatcaag 1560
atcttctctg aagttgaacc aaatccatcc accaactctg tctacaaggg tttggaaatg 1620
atggtcgatt tccaaccaga caccatcatt gctttgggtg gtggttctgc catggatgct 1680
gccaaggcta tgtggatggt ctctgaacat ccagaaactt ctttcttcgg tgccaagcaa 1740
aaattcttgg acattggtaa gagaacctac aagattggta tgccagaaaa cgccactttc 1800
atctgtattc caaccacttc tgggtactgg tctgaagtca ctccatttgc tgttatcact 1860
gactctgaaa ccaacgtcaa ataccatttg gctgatttcg ctttgactcc agatgtcgcc 1920
atcattgacc ctcaatttgt catgtccgtc ccaaaatctg tcaactgctga taccggtatg 1980
gacgttttga ctcacgctat ggaatcttac gtttctgtca tggcctccga ttacaccaga 2040
ggtttgtccc tacaagctat caaattgacc tttgaatact tgaaatcttc cgttgaaaaa 2100
ggtgacaagg tttccagaga aaagatgcac aacgcttcta ctttggccgg tatggccttt 2160
gctaacgctt tcttgggtat tgcctactcc attgctcaca aaattggtgg tgaatacggc 2220
attccacatg gtagagctaa cgccatcttg ttgcctcaca tcatcagata caacgccaag 2280
gaccctcaaa agcacgcttt gttcccaaag tacgaattct tcagagctga caccgattac 2340
gctgatatcg ccaagttctt aggtttgaaa ggtaacacca ctgaagcttt ggttgaatct 2400
ttggccaagg ctgtctacga attaggtcaa tctgttggtg ttgaaatgaa cttgaaatct 2460
caaggtgtct ctgaagaaga attgaacgaa tccattgaca gaatggctga attggctttc 2520
gaagaccaat gtaccactgc caacccaaag gaagctttga tttctgaaat caaggatatc 2580
atccaaactt cttacgacta caagcag 2607

```

<210> SEQ ID NO 30

<211> LENGTH: 392

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 30

```

Met Lys Glu Val Val Ile Ala Ser Ala Val Arg Thr Ala Ile Gly Ser
1           5           10          15

```

```

Tyr Gly Lys Ser Leu Lys Asp Val Pro Ala Val Asp Leu Gly Ala Thr
20          25          30

```

```

Ala Ile Lys Glu Ala Val Lys Lys Ala Gly Ile Lys Pro Glu Asp Val
35          40          45

```

```

Asn Glu Val Ile Leu Gly Asn Val Leu Gln Ala Gly Leu Gly Gln Asn
50          55          60

```

```

Pro Ala Arg Gln Ala Ser Phe Lys Ala Gly Leu Pro Val Glu Ile Pro

```

-continued

65	70	75	80
Ala Met Thr Ile Asn Lys Val Cys Gly Ser Gly Leu Arg Thr Val Ser	85	90	95
Leu Ala Ala Gln Ile Ile Lys Ala Gly Asp Ala Asp Val Ile Ile Ala	100	105	110
Gly Gly Met Glu Asn Met Ser Arg Ala Pro Tyr Leu Ala Asn Asn Ala	115	120	125
Arg Trp Gly Tyr Arg Met Gly Asn Ala Lys Phe Val Asp Glu Met Ile	130	135	140
Thr Asp Gly Leu Trp Asp Ala Phe Asn Asp Tyr His Met Gly Ile Thr	145	150	155
Ala Glu Asn Ile Ala Glu Arg Trp Asn Ile Ser Arg Glu Glu Gln Asp	165	170	175
Glu Phe Ala Leu Ala Ser Gln Lys Lys Ala Glu Glu Ala Ile Lys Ser	180	185	190
Gly Gln Phe Lys Asp Glu Ile Val Pro Val Val Ile Lys Gly Arg Lys	195	200	205
Gly Glu Thr Val Val Asp Thr Asp Glu His Pro Arg Phe Gly Ser Thr	210	215	220
Ile Glu Gly Leu Ala Lys Leu Lys Pro Ala Phe Lys Lys Asp Gly Thr	225	230	235
Val Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Cys Ala Ala Val Leu	245	250	255
Val Ile Met Ser Ala Glu Lys Ala Lys Glu Leu Gly Val Lys Pro Leu	260	265	270
Ala Lys Ile Val Ser Tyr Gly Ser Ala Gly Val Asp Pro Ala Ile Met	275	280	285
Gly Tyr Gly Pro Phe Tyr Ala Thr Lys Ala Ala Ile Glu Lys Ala Gly	290	295	300
Trp Thr Val Asp Glu Leu Asp Leu Ile Glu Ser Asn Glu Ala Phe Ala	305	310	315
Ala Gln Ser Leu Ala Val Ala Lys Asp Leu Lys Phe Asp Met Asn Lys	325	330	335
Val Asn Val Asn Gly Gly Ala Ile Ala Leu Gly His Pro Ile Gly Ala	340	345	350
Ser Gly Ala Arg Ile Leu Val Thr Leu Val His Ala Met Gln Lys Arg	355	360	365
Asp Ala Lys Lys Gly Leu Ala Thr Leu Cys Ile Gly Gly Gly Gln Gly	370	375	380
Thr Ala Ile Leu Leu Glu Lys Cys	385	390	

<210> SEQ ID NO 31

<211> LENGTH: 282

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 31

Met Lys Lys Val Cys Val Ile Gly Ala Gly Thr Met Gly Ser Gly Ile	1	5	10	15
Ala Gln Ala Phe Ala Ala Lys Gly Phe Glu Val Val Leu Arg Asp Ile	20	25	30	

-continued

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

<210> SEQ ID NO 32

<211> LENGTH: 261

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 32

Met Glu Leu Asn Asn Val Ile Leu Glu Lys Glu Gly Lys Val Ala Val
 1 5 10 15
 Val Thr Ile Asn Arg Pro Lys Ala Leu Asn Ala Leu Asn Ser Asp Thr
 20 25 30
 Leu Lys Glu Met Asp Tyr Val Ile Gly Glu Ile Glu Asn Asp Ser Glu
 35 40 45
 Val Leu Ala Val Ile Leu Thr Gly Ala Gly Glu Lys Ser Phe Val Ala
 50 55 60
 Gly Ala Asp Ile Ser Glu Met Lys Glu Met Asn Thr Ile Glu Gly Arg
 65 70 75 80
 Lys Phe Gly Ile Leu Gly Asn Lys Val Phe Arg Arg Leu Glu Leu Leu
 85 90 95
 Glu Lys Pro Val Ile Ala Ala Val Asn Gly Phe Ala Leu Gly Gly Gly
 100 105 110

-continued

Cys Glu Ile Ala Met Ser Cys Asp Ile Arg Ile Ala Ser Ser Asn Ala
 115 120 125
 Arg Phe Gly Gln Pro Glu Val Gly Leu Gly Ile Thr Pro Gly Phe Gly
 130 135 140
 Gly Thr Gln Arg Leu Ser Arg Leu Val Gly Met Gly Met Ala Lys Gln
 145 150 155 160
 Leu Ile Phe Thr Ala Gln Asn Ile Lys Ala Asp Glu Ala Leu Arg Ile
 165 170 175
 Gly Leu Val Asn Lys Val Val Glu Pro Ser Glu Leu Met Asn Thr Ala
 180 185 190
 Lys Glu Ile Ala Asn Lys Ile Val Ser Asn Ala Pro Val Ala Val Lys
 195 200 205
 Leu Ser Lys Gln Ala Ile Asn Arg Gly Met Gln Cys Asp Ile Asp Thr
 210 215 220
 Ala Leu Ala Phe Glu Ser Glu Ala Phe Gly Glu Cys Phe Ser Thr Glu
 225 230 235 240
 Asp Gln Lys Asp Ala Met Thr Ala Phe Ile Glu Lys Arg Lys Ile Glu
 245 250 255
 Gly Phe Lys Asn Arg
 260

<210> SEQ ID NO 33

<211> LENGTH: 379

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 33

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

-continued

195					200					205					
Thr	Thr	Glu	Leu	Val	Phe	Glu	Asp	Met	Ile	Val	Pro	Val	Glu	Asn	Met
210					215					220					
Ile	Gly	Lys	Glu	Gly	Lys	Gly	Phe	Pro	Ile	Ala	Met	Lys	Thr	Leu	Asp
225					230					235					240
Gly	Gly	Arg	Ile	Gly	Ile	Ala	Ala	Gln	Ala	Leu	Gly	Ile	Ala	Glu	Gly
				245					250					255	
Ala	Phe	Asn	Glu	Ala	Arg	Ala	Tyr	Met	Lys	Glu	Arg	Lys	Gln	Phe	Gly
			260					265					270		
Arg	Ser	Leu	Asp	Lys	Phe	Gln	Gly	Leu	Ala	Trp	Met	Met	Ala	Asp	Met
			275				280					285			
Asp	Val	Ala	Ile	Glu	Ser	Ala	Arg	Tyr	Leu	Val	Tyr	Lys	Ala	Ala	Tyr
			290				295					300			
Leu	Lys	Gln	Ala	Gly	Leu	Pro	Tyr	Thr	Val	Asp	Ala	Ala	Arg	Ala	Lys
305					310					315					320
Leu	His	Ala	Ala	Asn	Val	Ala	Met	Asp	Val	Thr	Thr	Lys	Ala	Val	Gln
				325					330					335	
Leu	Phe	Gly	Gly	Tyr	Gly	Tyr	Thr	Lys	Asp	Tyr	Pro	Val	Glu	Arg	Met
			340					345					350		
Met	Arg	Asp	Ala	Lys	Ile	Thr	Glu	Ile	Tyr	Glu	Gly	Thr	Ser	Glu	Val
			355				360					365			
Gln	Lys	Leu	Val	Ile	Ser	Gly	Lys	Ile	Phe	Arg					
			370				375								

<210> SEQ ID NO 34

<211> LENGTH: 858

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 34

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

Leu	Ser	Gln	Asp	Leu	Met	Ser	Glu	Ala	Asp	Ile	Ile	Leu	Ala	Thr	Gly
			180				185						190		
Gly	Pro	Ser	Met	Val	Lys	Ala	Ala	Tyr	Ser	Ser	Gly	Lys	Pro	Ala	Ile
			195				200						205		
Gly	Val	Gly	Ala	Gly	Asn	Thr	Pro	Ala	Ile	Ile	Asp	Glu	Ser	Ala	Asp
			210				215						220		
Ile	Asp	Met	Ala	Val	Ser	Ser	Ile	Ile	Leu	Ser	Lys	Thr	Tyr	Asp	Asn
			225				230						235		
Gly	Val	Ile	Cys	Ala	Ser	Glu	Gln	Ser	Ile	Leu	Val	Met	Asn	Ser	Ile
			245				250						255		
Tyr	Glu	Lys	Val	Lys	Glu	Glu	Phe	Val	Lys	Arg	Gly	Ser	Tyr	Ile	Leu
			260				265						270		
Asn	Gln	Asn	Glu	Ile	Ala	Lys	Ile	Lys	Glu	Thr	Met	Phe	Lys	Asn	Gly
			275				280						285		
Ala	Ile	Asn	Ala	Asp	Ile	Val	Gly	Lys	Ser	Ala	Tyr	Ile	Ile	Ala	Lys
			290				295						300		
Met	Ala	Gly	Ile	Glu	Val	Pro	Gln	Thr	Thr	Lys	Ile	Leu	Ile	Gly	Glu
			305				310						315		
Val	Gln	Ser	Val	Glu	Lys	Ser	Glu	Leu	Phe	Ser	His	Glu	Lys	Leu	Ser
			325				330						335		
Pro	Val	Leu	Ala	Met	Tyr	Lys	Val	Lys	Asp	Phe	Asp	Glu	Ala	Leu	Lys
			340				345						350		
Lys	Ala	Gln	Arg	Leu	Ile	Glu	Leu	Gly	Gly	Ser	Gly	His	Thr	Ser	Ser
			355				360						365		
Leu	Tyr	Ile	Asp	Ser	Gln	Asn	Asn	Lys	Asp	Lys	Val	Lys	Glu	Phe	Gly
			370				375						380		
Leu	Ala	Met	Lys	Thr	Ser	Arg	Thr	Phe	Ile	Asn	Met	Pro	Ser	Ser	Gln
			385				390						395		
Gly	Ala	Ser	Gly	Asp	Leu	Tyr	Asn	Phe	Ala	Ile	Ala	Pro	Ser	Phe	Thr
			405				410						415		
Leu	Gly	Cys	Gly	Thr	Trp	Gly	Gly	Asn	Ser	Val	Ser	Gln	Asn	Val	Glu
			420				425						430		
Pro	Lys	His	Leu	Leu	Asn	Ile	Lys	Ser	Val	Ala	Glu	Arg	Arg	Glu	Asn
			435				440						445		
Met	Leu	Trp	Phe	Lys	Val	Pro	Gln	Lys	Ile	Tyr	Phe	Lys	Tyr	Gly	Cys
			450				455						460		
Leu	Arg	Phe	Ala	Leu	Lys	Glu	Leu	Lys	Asp	Met	Asn	Lys	Lys	Arg	Ala
			465				470						475		
Phe	Ile	Val	Thr	Asp	Lys	Asp	Leu	Phe	Lys	Leu	Gly	Tyr	Val	Asn	Lys
			485				490						495		
Ile	Thr	Lys	Val	Leu	Asp	Glu	Ile	Asp	Ile	Lys	Tyr	Ser	Ile	Phe	Thr
			500				505						510		
Asp	Ile	Lys	Ser	Asp	Pro	Thr	Ile	Asp	Ser	Val	Lys	Lys	Gly	Ala	Lys
			515				520						525		
Glu	Met	Leu	Asn	Phe	Glu	Pro	Asp	Thr	Ile	Ile	Ser	Ile	Gly	Gly	Gly
			530				535						540		
Ser	Pro	Met	Asp	Ala	Ala	Lys	Val	Met	His	Leu	Leu	Tyr	Glu	Tyr	Pro
			545				550						555		
Glu	Ala	Glu	Ile	Glu	Asn	Leu	Ala	Ile	Asn	Phe	Met	Asp	Ile	Arg	Lys
			565				570						575		
Arg	Ile	Cys	Asn	Phe	Pro	Lys	Leu	Gly	Thr	Lys	Ala	Ile	Ser	Val	Ala

-continued

580					585					590						
Ile	Pro	Thr	Thr	Ala	Gly	Thr	Gly	Ser	Glu	Ala	Thr	Pro	Phe	Ala	Val	
595					600					605						
Ile	Thr	Asn	Asp	Glu	Thr	Gly	Met	Lys	Tyr	Pro	Leu	Thr	Ser	Tyr	Glu	
610					615					620						
Leu	Thr	Pro	Asn	Met	Ala	Ile	Ile	Asp	Thr	Glu	Leu	Met	Leu	Asn	Met	
625					630					635					640	
Pro	Arg	Lys	Leu	Thr	Ala	Ala	Thr	Gly	Ile	Asp	Ala	Leu	Val	His	Ala	
645					650					655						
Ile	Glu	Ala	Tyr	Val	Ser	Val	Met	Ala	Thr	Asp	Tyr	Thr	Asp	Glu	Leu	
660					665					670						
Ala	Leu	Arg	Ala	Ile	Lys	Met	Ile	Phe	Lys	Tyr	Leu	Pro	Arg	Ala	Tyr	
675					680					685						
Lys	Asn	Gly	Thr	Asn	Asp	Ile	Glu	Ala	Arg	Glu	Lys	Met	Ala	His	Ala	
690					695					700						
Ser	Asn	Ile	Ala	Gly	Met	Ala	Phe	Ala	Asn	Ala	Phe	Leu	Gly	Val	Cys	
705					710					715					720	
His	Ser	Met	Ala	His	Lys	Leu	Gly	Ala	Met	His	His	Val	Pro	His	Gly	
725					730					735						
Ile	Ala	Cys	Ala	Val	Leu	Ile	Glu	Glu	Val	Ile	Lys	Tyr	Asn	Ala	Thr	
740					745					750						
Asp	Cys	Pro	Thr	Lys	Gln	Thr	Ala	Phe	Pro	Gln	Tyr	Lys	Ser	Pro	Asn	
755					760					765						
Ala	Lys	Arg	Lys	Tyr	Ala	Glu	Ile	Ala	Glu	Tyr	Leu	Asn	Leu	Lys	Gly	
770					775					780						
Thr	Ser	Asp	Thr	Glu	Lys	Val	Thr	Ala	Leu	Ile	Glu	Ala	Ile	Ser	Lys	
785					790					795					800	
Leu	Lys	Ile	Asp	Leu	Ser	Ile	Pro	Gln	Asn	Ile	Ser	Ala	Ala	Gly	Ile	
805					810					815						
Asn	Lys	Lys	Asp	Phe	Tyr	Asn	Thr	Leu	Asp	Lys	Met	Ser	Glu	Leu	Ala	
820					825					830						
Phe	Asp	Asp	Gln	Cys	Thr	Thr	Ala	Asn	Pro	Arg	Tyr	Pro	Leu	Ile	Ser	
835					840					845						
Glu	Leu	Lys	Asp	Ile	Tyr	Ile	Lys	Ser	Phe							
850					855											

<210> SEQ ID NO 35

<211> LENGTH: 862

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 35

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

-continued

Lys	Asp	Glu	Lys	Thr	Cys	Gly	Ile	Ile	Glu	Arg	Asn	Glu	Pro	Tyr	Gly
			85						90					95	
Ile	Thr	Lys	Ile	Ala	Glu	Pro	Ile	Gly	Val	Val	Ala	Ala	Ile	Ile	Pro
			100					105					110		
Val	Thr	Asn	Pro	Thr	Ser	Thr	Thr	Ile	Phe	Lys	Ser	Leu	Ile	Ser	Leu
		115					120					125			
Lys	Thr	Arg	Asn	Gly	Ile	Phe	Phe	Ser	Pro	His	Pro	Arg	Ala	Lys	Lys
	130					135					140				
Ser	Thr	Ile	Leu	Ala	Ala	Lys	Thr	Ile	Leu	Asp	Ala	Ala	Val	Lys	Ser
145					150					155				160	
Gly	Ala	Pro	Glu	Asn	Ile	Ile	Gly	Trp	Ile	Asp	Glu	Pro	Ser	Ile	Glu
			165					170						175	
Leu	Thr	Gln	Tyr	Leu	Met	Gln	Lys	Ala	Asp	Ile	Thr	Leu	Ala	Thr	Gly
		180					185						190		
Gly	Pro	Ser	Leu	Val	Lys	Ser	Ala	Tyr	Ser	Ser	Gly	Lys	Pro	Ala	Ile
	195						200					205			
Gly	Val	Gly	Pro	Gly	Asn	Thr	Pro	Val	Ile	Ile	Asp	Glu	Ser	Ala	His
	210					215					220				
Ile	Lys	Met	Ala	Val	Ser	Ser	Ile	Ile	Leu	Ser	Lys	Thr	Tyr	Asp	Asn
225					230					235				240	
Gly	Val	Ile	Cys	Ala	Ser	Glu	Gln	Ser	Val	Ile	Val	Leu	Lys	Ser	Ile
			245					250						255	
Tyr	Asn	Lys	Val	Lys	Asp	Glu	Phe	Gln	Glu	Arg	Gly	Ala	Tyr	Ile	Ile
	260						265						270		
Lys	Lys	Asn	Glu	Leu	Asp	Lys	Val	Arg	Glu	Val	Ile	Phe	Lys	Asp	Gly
	275						280					285			
Ser	Val	Asn	Pro	Lys	Ile	Val	Gly	Gln	Ser	Ala	Tyr	Thr	Ile	Ala	Ala
	290					295				300					
Met	Ala	Gly	Ile	Lys	Val	Pro	Lys	Thr	Thr	Arg	Ile	Leu	Ile	Gly	Glu
305					310					315				320	
Val	Thr	Ser	Leu	Gly	Glu	Glu	Glu	Pro	Phe	Ala	His	Glu	Lys	Leu	Ser
			325					330						335	
Pro	Val	Leu	Ala	Met	Tyr	Glu	Ala	Asp	Asn	Phe	Asp	Asp	Ala	Leu	Lys
		340					345						350		
Lys	Ala	Val	Thr	Leu	Ile	Asn	Leu	Gly	Gly	Leu	Gly	His	Thr	Ser	Gly
	355					360						365			
Ile	Tyr	Ala	Asp	Glu	Ile	Lys	Ala	Arg	Asp	Lys	Ile	Asp	Arg	Phe	Ser
	370					375				380					
Ser	Ala	Met	Lys	Thr	Val	Arg	Thr	Phe	Val	Asn	Ile	Pro	Thr	Ser	Gln
385					390					395				400	
Gly	Ala	Ser	Gly	Asp	Leu	Tyr	Asn	Phe	Arg	Ile	Pro	Pro	Ser	Phe	Thr
			405					410						415	
Leu	Gly	Cys	Gly	Phe	Trp	Gly	Gly	Asn	Ser	Val	Ser	Glu	Asn	Val	Gly
		420					425						430		
Pro	Lys	His	Leu	Leu	Asn	Ile	Lys	Thr	Val	Ala	Glu	Arg	Arg	Glu	Asn
		435				440						445			
Met	Leu	Trp	Phe	Arg	Val	Pro	His	Lys	Val	Tyr	Phe	Lys	Phe	Gly	Cys
	450					455				460					
Leu	Gln	Phe	Ala	Leu	Lys	Asp	Leu	Lys	Asp	Leu	Lys	Lys	Lys	Arg	Ala
465					470					475				480	
Phe	Ile	Val	Thr	Asp	Ser	Asp	Pro	Tyr	Asn	Leu	Asn	Tyr	Val	Asp	Ser

-continued

485					490					495					
Ile	Ile	Lys	Ile	Leu	Glu	His	Leu	Asp	Ile	Asp	Phe	Lys	Val	Phe	Asn
			500					505					510		
Lys	Val	Gly	Arg	Glu	Ala	Asp	Leu	Lys	Thr	Ile	Lys	Lys	Ala	Thr	Glu
		515					520					525			
Glu	Met	Ser	Ser	Phe	Met	Pro	Asp	Thr	Ile	Ile	Ala	Leu	Gly	Gly	Thr
	530					535					540				
Pro	Glu	Met	Ser	Ser	Ala	Lys	Leu	Met	Trp	Val	Leu	Tyr	Glu	His	Pro
545					550					555					560
Glu	Val	Lys	Phe	Glu	Asp	Leu	Ala	Ile	Lys	Phe	Met	Asp	Ile	Arg	Lys
				565					570					575	
Arg	Ile	Tyr	Thr	Phe	Pro	Lys	Leu	Gly	Lys	Lys	Ala	Met	Leu	Val	Ala
			580					585						590	
Ile	Thr	Thr	Ser	Ala	Gly	Ser	Gly	Ser	Glu	Val	Thr	Pro	Phe	Ala	Leu
			595				600					605			
Val	Thr	Asp	Asn	Asn	Thr	Gly	Asn	Lys	Tyr	Met	Leu	Ala	Asp	Tyr	Glu
	610					615					620				
Met	Thr	Pro	Asn	Met	Ala	Ile	Val	Asp	Ala	Glu	Leu	Met	Met	Lys	Met
625					630					635					640
Pro	Lys	Gly	Leu	Thr	Ala	Tyr	Ser	Gly	Ile	Asp	Ala	Leu	Val	Asn	Ser
				645					650					655	
Ile	Glu	Ala	Tyr	Thr	Ser	Val	Tyr	Ala	Ser	Glu	Tyr	Thr	Asn	Gly	Leu
			660					665					670		
Ala	Leu	Glu	Ala	Ile	Arg	Leu	Ile	Phe	Lys	Tyr	Leu	Pro	Glu	Ala	Tyr
		675					680					685			
Lys	Asn	Gly	Arg	Thr	Asn	Glu	Lys	Ala	Arg	Glu	Lys	Met	Ala	His	Ala
	690					695					700				
Ser	Thr	Met	Ala	Gly	Met	Ala	Ser	Ala	Asn	Ala	Phe	Leu	Gly	Leu	Cys
705					710					715					720
His	Ser	Met	Ala	Ile	Lys	Leu	Ser	Ser	Glu	His	Asn	Ile	Pro	Ser	Gly
				725					730					735	
Ile	Ala	Asn	Ala	Leu	Leu	Ile	Glu	Glu	Val	Ile	Lys	Phe	Asn	Ala	Val
			740				745						750		
Asp	Asn	Pro	Val	Lys	Gln	Ala	Pro	Cys	Pro	Gln	Tyr	Lys	Tyr	Pro	Asn
		755					760					765			
Thr	Ile	Phe	Arg	Tyr	Ala	Arg	Ile	Ala	Asp	Tyr	Ile	Lys	Leu	Gly	Gly
	770					775					780				
Asn	Thr	Asp	Glu	Glu	Lys	Val	Asp	Leu	Leu	Ile	Asn	Lys	Ile	His	Glu
785					790					795					800
Leu	Lys	Lys	Ala	Leu	Asn	Ile	Pro	Thr	Ser	Ile	Lys	Asp	Ala	Gly	Val
				805					810					815	
Leu	Glu	Glu	Asn	Phe	Tyr	Ser	Ser	Leu	Asp	Arg	Ile	Ser	Glu	Leu	Ala
			820					825					830		
Leu	Asp	Asp	Gln	Cys	Thr	Gly	Ala	Asn	Pro	Arg	Phe	Pro	Leu	Thr	Ser
		835					840					845			
Glu	Ile	Lys	Glu	Met	Tyr	Ile	Asn	Cys	Phe	Lys	Lys	Gln	Pro		
	850					855					860				

<210> SEQ ID NO 36

<211> LENGTH: 389

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

-continued

<400> SEQUENCE: 36

```

Met Leu Ser Phe Asp Tyr Ser Ile Pro Thr Lys Val Phe Phe Gly Lys
 1           5           10           15
Gly Lys Ile Asp Val Ile Gly Glu Glu Ile Lys Lys Tyr Gly Ser Arg
          20           25           30
Val Leu Ile Val Tyr Gly Gly Gly Ser Ile Lys Arg Asn Gly Ile Tyr
          35           40           45
Asp Arg Ala Thr Ala Ile Leu Lys Glu Asn Asn Ile Ala Phe Tyr Glu
          50           55           60
Leu Ser Gly Val Glu Pro Asn Pro Arg Ile Thr Thr Val Lys Lys Gly
          65           70           75           80
Ile Glu Ile Cys Arg Glu Asn Asn Val Asp Leu Val Leu Ala Ile Gly
          85           90           95
Gly Gly Ser Ala Ile Asp Cys Ser Lys Val Ile Ala Ala Gly Val Tyr
          100          105          110
Tyr Asp Gly Asp Thr Trp Asp Met Val Lys Asp Pro Ser Lys Ile Thr
          115          120          125
Lys Val Leu Pro Ile Ala Ser Ile Leu Thr Leu Ser Ala Thr Gly Ser
          130          135          140
Glu Met Asp Gln Ile Ala Val Ile Ser Asn Met Glu Thr Asn Glu Lys
          145          150          155          160
Leu Gly Val Gly His Asp Asp Met Arg Pro Lys Phe Ser Val Leu Asp
          165          170          175
Pro Thr Tyr Thr Phe Thr Val Pro Lys Asn Gln Thr Ala Ala Gly Thr
          180          185          190
Ala Asp Ile Met Ser His Thr Phe Glu Ser Tyr Phe Ser Gly Val Glu
          195          200          205
Gly Ala Tyr Val Gln Asp Gly Ile Ala Glu Ala Ile Leu Arg Thr Cys
          210          215          220
Ile Lys Tyr Gly Lys Ile Ala Met Glu Lys Thr Asp Asp Tyr Glu Ala
          225          230          235          240
Arg Ala Asn Leu Met Trp Ala Ser Ser Leu Ala Ile Asn Gly Leu Leu
          245          250          255
Ser Leu Gly Lys Asp Arg Lys Trp Ser Cys His Pro Met Glu His Glu
          260          265          270
Leu Ser Ala Tyr Tyr Asp Ile Thr His Gly Val Gly Leu Ala Ile Leu
          275          280          285
Thr Pro Asn Trp Met Glu Tyr Ile Leu Asn Asp Asp Thr Leu His Lys
          290          295          300
Phe Val Ser Tyr Gly Ile Asn Val Trp Gly Ile Asp Lys Asn Lys Asp
          305          310          315          320
Asn Tyr Glu Ile Ala Arg Glu Ala Ile Lys Asn Thr Arg Glu Tyr Phe
          325          330          335
Asn Ser Leu Gly Ile Pro Ser Lys Leu Arg Glu Val Gly Ile Gly Lys
          340          345          350
Asp Lys Leu Glu Leu Met Ala Lys Gln Ala Val Arg Asn Ser Gly Gly
          355          360          365
Thr Ile Gly Ser Leu Arg Pro Ile Asn Ala Glu Asp Val Leu Glu Ile
          370          375          380
Phe Lys Lys Ser Tyr

```

-continued

385

<210> SEQ ID NO 37

<211> LENGTH: 390

<212> TYPE: PRT

<213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 37

Met Val Asp Phe Glu Tyr Ser Ile Pro Thr Arg Ile Phe Phe Gly Lys
1 5 10 15
Asp Lys Ile Asn Val Leu Gly Arg Glu Leu Lys Lys Tyr Gly Ser Lys
20 25 30
Val Leu Ile Val Tyr Gly Gly Glu Ser Ile Lys Arg Asn Gly Ile Tyr
35 40 45
Asp Lys Ala Val Ser Ile Leu Glu Lys Asn Ser Ile Lys Phe Tyr Glu
50 55 60
Leu Ala Gly Val Glu Pro Asn Pro Arg Val Thr Thr Val Glu Lys Gly
65 70 75 80
Val Lys Ile Cys Arg Glu Asn Gly Val Glu Val Val Leu Ala Ile Gly
85 90 95
Gly Gly Ser Ala Ile Asp Cys Ala Lys Val Ile Ala Ala Ala Cys Glu
100 105 110
Tyr Asp Gly Asn Pro Trp Asp Ile Val Leu Asp Gly Ser Lys Ile Lys
115 120 125
Arg Val Leu Pro Ile Ala Ser Ile Leu Thr Ile Ala Ala Thr Gly Ser
130 135 140
Glu Met Asp Thr Trp Ala Val Ile Asn Asn Met Asp Thr Asn Glu Lys
145 150 155 160
Leu Ile Ala Ala His Pro Asp Met Ala Pro Lys Phe Ser Ile Leu Asp
165 170 175
Pro Thr Tyr Thr Tyr Thr Val Pro Thr Asn Gln Thr Ala Ala Gly Thr
180 185 190
Ala Asp Ile Met Ser His Ile Phe Glu Val Tyr Phe Ser Asn Thr Lys
195 200 205
Thr Ala Tyr Leu Gln Asp Arg Met Ala Glu Ala Leu Leu Arg Thr Cys
210 215 220
Ile Lys Tyr Gly Gly Ile Ala Leu Glu Lys Pro Asp Asp Tyr Glu Ala
225 230 235 240
Arg Ala Asn Leu Met Trp Ala Ser Ser Leu Ala Ile Asn Gly Leu Leu
245 250 255
Thr Tyr Gly Lys Asp Thr Asn Trp Ser Val His Leu Met Glu His Glu
260 265 270
Leu Ser Ala Tyr Tyr Asp Ile Thr His Gly Val Gly Leu Ala Ile Leu
275 280 285
Thr Pro Asn Trp Met Glu Tyr Ile Leu Asn Asn Asp Thr Val Tyr Lys
290 295 300
Phe Val Glu Tyr Gly Val Asn Val Trp Gly Ile Asp Lys Glu Lys Asn
305 310 315 320
His Tyr Asp Ile Ala His Gln Ala Ile Gln Lys Thr Arg Asp Tyr Phe
325 330 335
Val Asn Val Leu Gly Leu Pro Ser Arg Leu Arg Asp Val Gly Ile Glu
340 345 350

-continued

Glu Glu Lys Leu Asp Ile Met Ala Lys Glu Ser Val Lys Leu Thr Gly
 355 360 365
 Gly Thr Ile Gly Asn Leu Arg Pro Val Asn Ala Ser Glu Val Leu Gln
 370 375 380
 Ile Phe Lys Lys Ser Val
 385 390
 <210> SEQ ID NO 38
 <211> LENGTH: 336
 <212> TYPE: PRT
 <213> ORGANISM: Clostridium acetobutylicum
 <400> SEQUENCE: 38
 Met Asn Lys Ala Asp Tyr Lys Gly Val Trp Val Phe Ala Glu Gln Arg
 1 5 10 15
 Asp Gly Glu Leu Gln Lys Val Ser Leu Glu Leu Leu Gly Lys Gly Lys
 20 25 30
 Glu Met Ala Glu Lys Leu Gly Val Glu Leu Thr Ala Val Leu Leu Gly
 35 40 45
 His Asn Thr Glu Lys Met Ser Lys Asp Leu Leu Ser His Gly Ala Asp
 50 55 60
 Lys Val Leu Ala Ala Asp Asn Glu Leu Leu Ala His Phe Ser Thr Asp
 65 70 75 80
 Gly Tyr Ala Lys Val Ile Cys Asp Leu Val Asn Glu Arg Lys Pro Glu
 85 90 95
 Ile Leu Phe Ile Gly Ala Thr Phe Ile Gly Arg Asp Leu Gly Pro Arg
 100 105 110
 Ile Ala Ala Arg Leu Ser Thr Gly Leu Thr Ala Asp Cys Thr Ser Leu
 115 120 125
 Asp Ile Asp Val Glu Asn Arg Asp Leu Leu Ala Thr Arg Pro Ala Phe
 130 135 140
 Gly Gly Asn Leu Ile Ala Thr Ile Val Cys Ser Asp His Arg Pro Gln
 145 150 155 160
 Met Ala Thr Val Arg Pro Gly Val Phe Glu Lys Leu Pro Val Asn Asp
 165 170 175
 Ala Asn Val Ser Asp Asp Lys Ile Glu Lys Val Ala Ile Lys Leu Thr
 180 185 190
 Ala Ser Asp Ile Arg Thr Lys Val Ser Lys Val Val Lys Leu Ala Lys
 195 200 205
 Asp Ile Ala Asp Ile Gly Glu Ala Lys Val Leu Val Ala Gly Gly Arg
 210 215 220
 Gly Val Gly Ser Lys Glu Asn Phe Glu Lys Leu Glu Glu Leu Ala Ser
 225 230 235 240
 Leu Leu Gly Gly Thr Ile Ala Ala Ser Arg Ala Ala Ile Glu Lys Glu
 245 250 255
 Trp Val Asp Lys Asp Leu Gln Val Gly Gln Thr Gly Lys Thr Val Arg
 260 265 270
 Pro Thr Leu Tyr Ile Ala Cys Gly Ile Ser Gly Ala Ile Gln His Leu
 275 280 285
 Ala Gly Met Gln Asp Ser Asp Tyr Ile Ile Ala Ile Asn Lys Asp Val
 290 295 300
 Glu Ala Pro Ile Met Lys Val Ala Asp Leu Ala Ile Val Gly Asp Val
 305 310 315 320

-continued

Asn Lys Val Val Pro Glu Leu Ile Ala Gln Val Lys Ala Ala Asn Asn
 325 330 335

<210> SEQ ID NO 39
 <211> LENGTH: 259
 <212> TYPE: PRT
 <213> ORGANISM: Clostridium acetobutylicum

<400> SEQUENCE: 39

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

<210> SEQ ID NO 40
 <211> LENGTH: 5976
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: plasmid YEplac112PtdhTadh

<400> SEQUENCE: 40

gccccggggga tccactagtt ctagaatccg tcgaaactaa gttctggtgt tttaaaacta 60
 aaaaaaagac taactataaa agtagaattt aagaagtta agaaatagat ttacagaatt 120

-continued

acaatcaata cctacogtct ttatatactt attagtcaag taggggaata atttcaggga	180
actggtttca accttttttt tcagcttttt ccaaatcaga gagagcagaa ggtaatagaa	240
gggtgaagaa aatgagatag atacatgcgt gggcaattg ccttgtgtca tcatttactc	300
caggcagggt gcactactcc attgaggttg tgcccgtttt ttgctgttt gtgcccctgt	360
tctctgtagt tgcgctaaga gaatggacct atgaactgat ggttggtgaa gaaaacaata	420
ttttggtgct gggattcttt ttttttctgg atgccagctt aaaaagcggg ctccattata	480
tttagtggat gccaggaata aactgttcac ccagacacct acgatgttat atattctgtg	540
taaccgccc cctatttttg gcattgtacg gttacagcag aattaaaagg ctaatttttt	600
gactaaataa agttaggaaa atcactacta ttaattattt acgtattctt tgaatggcg	660
agtattgata atgataaact gagctcgaat tcaactggccg tcgttttaca acgtcgtgac	720
tgggaaaacc ctggcggttac ccaacttaat cgccttgacg cacatcccc ttctgccagc	780
tggcgtaata gcgaagaggc ccgcaccgat cgccttccc aacagttgag cagcctgaat	840
ggcgaatggc gcctgatgag gtattttctc cttacgcacg tgtgcggtat ttcacaccgc	900
atatatcgga tcgtacttgt taccatcat tgaattttga acatccgaac ctgggagttt	960
tccttgaaac agatagtata ttgaacctg tataataata tatagtctag cgctttacgg	1020
aagacaatgt atgtatttcg gttcctggag aaactattgc atctattgca taggtaatct	1080
tgcacgtcgc atccccggtt cattttctgc gtttccatct tgcacttcaa tagcatatct	1140
ttgttaacga agcatctgtg cttcattttg tagaacaaaa atgcaacgag agagcgctaa	1200
tttttcaaac aaagaatctg agctgcattt ttacagaaca gaaatgcaac gcgaaagcgc	1260
tattttacca acgaagaatc tgtgttcat ttttgtaaaa caaaaatgca acgcgagagc	1320
gctaattttt caaacaaaga atctgagctg catttttaca gaacagaaat gcaacgcgag	1380
agcgctattt taccaacaaa gaatctatac ttcttttttg ttctacaaaa atgcatcccg	1440
agagcgctat ttttctaaca aagcatctta gattactttt tttctccttt gtgcgctcta	1500
taatgcagtc tcttgataac tttttgact gtaggtcctg taaggttaga agaaggctac	1560
tttggtgtct attttctctt ccataaaaa agcctgactc cacttccgcg gtttactgat	1620
tactagcgaa gctgcgggtg cattttttca agataaaggc atccccgatt atattctata	1680
ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata gcgttgatga ttcttcattg	1740
gtcagaaaaat tatgaacggt ttcttctatt ttgtctctat atactacgta taggaaatgt	1800
ttacattttc gtattgtttt cgattcactc tatgaatagt tcttactaca atttttttgt	1860
ctaaagagta atactagaga taaacataaa aaatgtagag gtcgagtta gatgcaagtt	1920
caaggagcga aagggtgatg ggtaggttat atagggatat agcacagaga tatatagcaa	1980
agagatactt ttgagcaatg tttgtggaag cggatttcgc aatatttttag tagctcgta	2040
cagtccgggt cgtttttggt tttttgaaag tgcgtcttca gacgctttt ggttttcaaa	2100
agcgctctga agttctata ctttctagct agagaatagg aacttcgga taggaacttc	2160
aaagcgtttc cgaaaacgag cgcttccgaa aatgcaacgc gagctgcgca catacagctc	2220
actgttcacg tcgcacctat atctgcgtgt tgctgtata tatatatata tgagaagaac	2280
ggcatagtgc gtgtttatgc ttaaatgcgt acttatatgc gtctatttat gtaggatgaa	2340
aggtagtcta gtacctcctg tgatattatc ccattccatg cgggggatcg tatgcttcct	2400

-continued

tcagcactac cctttagctg ttctatatgc tgccactcct caattggatt agtctcatcc	2460
ttcaatgcta tcatttcctt tgatattgga tcgatccgat gataagctgt caaacatgag	2520
aattgatcctt ttatgcttgc ttttcaaaag gcttgccaggc aagtgcacaa acaatactta	2580
aataaatact actcagtaat aacctatttc ttagcatttt tgacgaaatt tgctattttg	2640
ttagagtctt ttacaccatt tgtctccaca cctccgctta catcaacacc aataacgcca	2700
tttaatctaa gcgcacaccc aacattttct ggcgtcagtc caccagctaa cataaaatgt	2760
aagctctcgg ggctctcttg ccttccaacc cagtcagaaa tcgagttcca atccaaaagt	2820
tcacctgtcc cacctgcttc tgaatcaaac aagggaataa acgaatgagg tttctgtgaa	2880
gctgcactga gtagtatgtt gcagtccttt ggaaatacga gtcttttaaat aactggcaaa	2940
ccgaggaaact cttggtatcc ttgccacgac tcactctccat gcagttggac gatatacatg	3000
ccgtaatcat tgaccagagc caaaacatcc tccttaggtt gattacgaaa cagccaacc	3060
aagtatttcg gagtgcctga actattttta tatgctttta caagacttga aattttcctt	3120
gcaataaacc ggtaaatgtt tctctttcta ttgggcacac atataatacc cagcaagtca	3180
gcatcggaat ctagtgcaca ttctgcggcc tctgtgctct gcaagccgca aactttcacc	3240
aatggaccag aactacctgt gaaattaata acagacatac tccaagctgc ctttgtgtgc	3300
ttaatcacgt atactcacgt gctcaatagt caccaatgcc ctccctcttg gccctctcct	3360
tttctttttt cgaccgaatt aattcttgaa gacgaaaggg cctcgtgata cgcctatttt	3420
tataggttaa tgatcatgata ataatggttt cttagacgtc aggtggcact tttcggggaa	3480
atgtgcgcgg aaccctattt tgtttatttt tctaaataca ttcaaatatg tatccgctca	3540
tgagacaata accctgataa atgcttcaat aatattgaaa aagggaagagt atgagtattc	3600
aacatttcgg tgtcgccctt attccctttt ttgcggcatt ttgccttccct gtttttgctc	3660
accagaaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt	3720
acatcgaaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacggt	3780
ttccaatgat gagcactttt aaagtctctg tatgtggcgc ggtattatcc cgtattgacg	3840
ccgggaaga gcaactcggt cgccgcatac actattctca gaatgacttg gttgagtact	3900
caccagtcac agaaaagcat cttacggatg gcacgacagt aagagaatta tgcagtgtctg	3960
ccataaccat gagtataac actgcggcca acttacttct gacaacgac ggaggaccga	4020
aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgctt gatcgttggg	4080
aaccggagct gaatgaagcc ataccaaagc acgagcgtga caccacgatg cctgtagcaa	4140
tggaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccgcaac	4200
aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc	4260
cggttggtctg gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca	4320
ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga	4380
gtcaggcaac tatggatgaa cgaaatagac agatcgctga gataggtgcc tcaactgatta	4440
agcattggta actgtcagac caagtttact catatatact ttagattgat ttaaaacttc	4500
atttttaatt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc	4560
cttaacgtga gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt	4620
cttgagatcc tttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac	4680

-continued

cagcgggtggt ttgtttgccg gatcaagagc taccaactct ttttccgaag gtaactggct	4740
tcagcagagc gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggcaccact	4800
tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg	4860
ctgccagtgg cgataagtgc tgtcttaccg ggttggaactc aagacgatag ttaccggata	4920
aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca gccagcttg gagcgaacga	4980
cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag	5040
ggagaaaggc ggacaggatc ccggttaagcg gcagggtcgg aacaggagag cgcacgagg	5100
agcttcacag gggaacgcc tggatcttt atagtcctgt cgggtttcgc cacctctgac	5160
ttgagcgtcg atttttgtga tgcctcgtcag gggggcggag cctatggaaa aacgccagca	5220
acgggcctt tttacggttc ctggcctttt gctggccttt tgcctacatg tttttctctg	5280
cgttatcccc tgattctgtg gataaccgta ttaccgcctt tgagttagct gataccgctc	5340
gccgcagccg aacgaccgag cgcagcgagt cagttagcga ggaagcggaa gagcgcccaa	5400
tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg cagcacagg	5460
ttcccgactg gaaagcgggc agtgagcgca acgcaattaa tgtgagttag ctactcatt	5520
aggcacccca ggctttacac tttatgcttc cggctcgtat gttgtgtgga attgtgagcg	5580
gataacaatt tcacacagga aacagctatg accatgatta cgccaagctt aggcctgtgt	5640
ggaagaacga ttacaacagg tgttgtcctc tgaggacata aaatacacac cgagattcat	5700
caactcattg ctggagttag catatctaca attgggtgaa atggggagcg atttgaggc	5760
atttgctcgg catgccggtg gaggtgtggt caataagagc gacctcatgc tatacctgag	5820
aaagcaacct gacctacagg aaagagttac tcaagaataa gaattttcgt tttaaaacct	5880
aagagtcact ttaaaatttg tatacactta ttttttttat aacttattha ataataaaaa	5940
tcataaatca taagaaatc gctcgagtcg actgca	5976

<210> SEQ ID NO 41

<211> LENGTH: 13286

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: pBOL34

<400> SEQUENCE: 41

aagettgcat gcctgcaggc cgacggcgcg cggggcccg ttaaacggcc ggccaagggtg	60
agacgcgcac aaccgctaga gtactttgaa gaggaacag caatagggtt gctaccagta	120
taaaagaca ggtacataca aactggaaa tggttgtctg tttgagtacg ctttcaattc	180
atttgggtgt gcactttatt atgttacaat atggaaggga actttacact tctctatgc	240
acatatatta attaaagtcc aatgctagta gagaaggggg gtaacacccc tccgcgtct	300
tttcgattt ttttctaaac cgtggaatat ttccgatatc cttttgttgt ttccgggtgt	360
acaatatgga cttcctcttt tctggcaacc aaaccatac atcgggattc ctataatacc	420
ttcgttggtc tccctaaccat gtaggtggcg gaggggagat atacaataga acagatacca	480
gacaagacat aatgggctaa acaagactac accaattaca ctgcctcatt gatgggtgta	540
cataacgaac taatactgta gccctagact tgatagccat catcatatcg aagtttact	600
accctttttc catttgccat ctattgaagt aataatagga gcatgcaact tcttttcttt	660

-continued

ttttttcttt tctctctccc cggttgttgt ctcaccatat cgcgaatgac aaaaaaatga	720
tggaagacac taaaggaaaa aattaacgac aaagacagca ccaacagatg tcgttgttcc	780
agagctgatg aggggtatct cgaagcacac gaaacttttt ccttccttca ttcacgcaca	840
ctactctcta atgagcaacg gtatacggcc ttcttccag ttacttgaat ttgaaataaa	900
aaaaagtttg ctgtcttgct atcaagtata aatagacctg caattattaa tcttttgttt	960
cctcgtcatt gttctcgttc cctttcttcc ttgtttcttt tctgcacaa tatttcaagc	1020
tataccaagc atacaatcaa ctatctcata tacaatgaag gaagttgtta ttgcttctgc	1080
tgtcagaact gccattggtt cttacggtaa gtctttgaag gacgtcccag ctgtcgactt	1140
gggtgctacc gccatcaagg aagctgtcaa gaaggctggt atcaagccag aagatgttaa	1200
cgaagttatc ttaggtaacg ttttgcaagc tggtttaggt caaaaccag ctcgtcaagc	1260
ttctttcaag gctggtttgc cagttgaaat tccagccatg accatcaaca aggtttgtgg	1320
ttctggtttg agaactgttt ctttggtgc tcaaatcatc aaggctggtg acgctgatgt	1380
catcattgct ggtggtatgg aaaacatgtc cagagctcca tacttggtta acaatgctag	1440
atggggttac agaatgggta acgccaagtt cgtcgatgaa atgatcactg acggtttatg	1500
ggacgctttc aacgactacc acatgggtat cactgctgaa aacattgctg aaagatggaa	1560
catctccaga gaagaacaag atgaatttgc tttggcttct caaagaagg ctgaagaagc	1620
catcaaatct ggtcaattca aggacgaaat tgtcccagtt gtcacaaagg gtagaaagg	1680
tgaaccggtt gtcgacaccg atgaacaccc aagattcggg tccaccattg aaggtttggc	1740
caagttgaaa ccagctttca agaaggatgg taccgctact gctggtaacg ctcccggttt	1800
gaacgactgt gctgctgttt tggttatcat gtctgctgaa aaggccaagg aattgggtgt	1860
caagccattg gccaagattg tctctacgg ttctgctggt gttgaccag ccatcatggg	1920
ttacggctct ttctacgcta ccaaggctgc tatcgaaaag gctggttga cgttgacga	1980
attggatttg attgaatcca acgaagcttt cgtgctcaa tctttggtg ttgccaagga	2040
cttgaaatc gatcatgaaca aggtcaacgt taacgggtgt gccattgctt tgggtcacc	2100
aattggtgct tccggtgcca gaatcttggt tacttttagt cacgctatgc aaaagcgtga	2160
tgccaagaag ggtttggtta ctctatgtat cgggtggtgt caaggctactg ccatcttatt	2220
ggaaaagtgt taggcccggt cataaagcaa tcttgatgag gataatgatt ttttttgaa	2280
tatacataaa tactaccgtt tttctgctag attttgtaa gacgtaaata agtacatatt	2340
actttttaag ccaagacaag attaagcatt aactttaccc tttctcttc taagtttcaa	2400
tactagtatt cactgtttaa aagttatggc gagaacgtcg gcggttaaaa tatattacc	2460
tgaacgtggt gaattgaagt tctaggatgg tttaaagatt tttcctttt gggaataag	2520
taacaatat attgctgcct ttgcaaacg cacataacca caatatgtga ctattggcaa	2580
agaacgcatt atcctttgaa gaggtggata ctgatactaa gagagtctct attccggctc	2640
cacttttagt ccagagatta cttgtcttct tacgtatcag aacaagaaag catttccaaa	2700
gtaattgcat ttgcccttga gcagtatata tatactaaga agtttaaaaca tttaaacgtg	2760
tgtgtgcatt atatatatta aaaattaaga attagactaa ataaagtgt tctaaaaaaa	2820
tattaaagtt gaaatgtgct tgttgtgaat tgtgctctat tagaataatt atgacttgtg	2880
tgcgtttcat attttaaaat aggaaataac caagaagaa aaagtaccat ccagagaaac	2940

-continued

caattatatac	aaatcaaata	aaacaaccag	cttcggtgtg	tgtgtgtgtg	tgaagctaag	3000
agttgatgcc	atttaatac	aaaatttta	ggtgtgtgtg	tggataaaat	attagaatga	3060
caattcgaga	tgaatttta	agcaaactct	agtaggaaat	aagcggctta	ttcttgttgg	3120
ctcctaattc	tttttagtgt	atcagttccc	attgataaaa	aaattaaaat	taaaattaga	3180
aaaattaaac	cagaaaaatc	aagttgatta	aatgtgaca	aaaattatga	ttaaatgcta	3240
cttcaacagg	agccccggcc	tatttggagt	agtcgtagaa	acccttacca	gactttctac	3300
ctaaccaacc	agctctaacg	tacttcttca	ataaagtgtg	aggctctgtac	ttagagtcac	3360
cggtttcaga	gtataagaca	tccatgatgg	ccaaacagat	atccaaaccg	atgaagtcac	3420
ctaattccaa	tggaccatt	gggtggttag	cacccaattt	catggccttg	tcgatatctt	3480
caacagaagc	aataccttca	gccaaaaatc	cgacagcttc	gttgatcatt	ggaatcaaga	3540
ttctgtgtac	aacgaaacct	ggagcttcag	caacttcaac	tgggtcctta	ccaatggcaa	3600
tggaggtttc	cttgacagca	tcgaaagttt	cttgagaggt	ggcaatacct	ctgatgactt	3660
cgaccaactt	catgactgga	gctgggttga	agaagtgcac	accgataacc	ttgtctggtc	3720
tcttggtagc	agaagcaact	tcagtgatgg	acaagaaga	agtgttgga	gccaaaatgg	3780
tttctggctt	acagatgttg	tccaaatcag	caaagatttg	cttcttgatg	tccattcttt	3840
caacggcagc	ttcaatgacc	aaatcacagt	cagcagccat	gttcaagtca	acagtaccgg	3900
agatttctgt	caagatttcg	accttggtag	cttcttcaat	cttacccttc	ttgaccaact	3960
tggacaagtt	cttggtgatg	aaatccaac	cacggtcaac	gaattcgtcc	ttgatatctc	4020
tcaaaacaac	ttcgaaaccc	ttggcagcga	aagcttgagc	aataccagaa	cccatggtac	4080
cggcaccaat	gacacaaacc	ttcttcattt	tgatttagtg	tttgtgtgtt	gataagcagt	4140
tgcttggttt	tttatgaaaa	atagctagaa	ggaataaggg	attacaagag	agatgttaca	4200
agaaagaagt	aaaataaatt	tgattaatat	tgccattatc	aaaagctatt	tatatgttga	4260
aatcgtggag	atcatgtgtg	ccagaaaagg	ccacagtttc	cggggagagg	cataccttga	4320
ggtggctagg	aatcacggag	acctcttgac	ttgcagggta	ggctagctag	aattaagtga	4380
ggtgacaagg	ttccataca	gttttgacct	tgagacgttg	ctacttacga	tttgagtat	4440
gcaagtctca	tgetgcaaac	aaaagaggac	cgctcaggta	atcgctcaat	tagtgagcgt	4500
tatcaggggc	gggagaggcg	aaagtgggtt	ttggtggtgt	aagtaaagg	cgcccaata	4560
tgcagggtgt	tgggtgctat	cctagtggaa	gctcgatca	gtagataacc	cgctagaag	4620
cggatatttt	cttttttttt	cttccttctt	tttcgtcatt	atttcaaacg	cttttgctgc	4680
aagtaatgaa	tatctggcgg	ttccgcggta	atgcgacaat	ttgtgatatg	cactcttaaa	4740
accccgccac	gatgatcgca	cgtgcggca	tttatagacg	acttttctgg	ttgtcccgct	4800
tcacggcaca	tgcattgcatc	aatgaccgaa	ttcagggttc	tactaaccat	tgtgttgtgt	4860
tattgtgtgt	catgaggtgc	tcaagtgc	gcgcatctg	actagtggta	actctagacg	4920
gcttcgatgc	agagagttcc	tcaaaatttt	tcttttcaat	tgtttgcctg	gtttccgcgg	4980
cgtatatcag	tttttgccga	tatggtaacg	cgatactcta	cggcaccttc	acggtagatg	5040
tcttttttaa	aagtgactgt	taattccagg	attgaaagga	agtgtcgaat	agtatagtat	5100
gctttctagg	ccggccggtt	aaatggggcc	gcgcccggtt	taaacggccg	gcccttcctt	5160
tttacagtgc	ttcgaaaaag	cacagcgttg	tccaaggga	caatttttct	tcaagttaat	5220

-continued

gcataagaaa	tatctttttt	tatgttttagc	taagtaaaag	cagcttggag	taaaaaaaa	5280
aatgagtaaa	tttctcgatg	gattagtttc	tcacaggtaa	cataacaaaa	accaagaaaa	5340
gcccgccttc	gaaaactaca	gttgacttgt	atgctaaagg	gccagactaa	tgggaggaga	5400
aaaaaaaacg	aatgtatatg	ctcattttaca	ctctatatca	ccatatggag	gataagttgg	5460
gctgagcttc	tgatccaatt	tattctatcc	attagtgtct	gatatgtccc	accagccaac	5520
acttgatagt	atctactcgc	cattcacttc	cagcagcgcc	agtagggttg	ttgagcttag	5580
taaaaatgtg	cgcaccacaa	gcctacatga	ctccacgtca	catgaaacca	caccgtgggg	5640
ccttggtgcg	ctaggaatag	gatatgcgac	gaagacgctt	ctgcttagta	accacaccac	5700
attttcaggg	ggtcgatctg	cttgcttctc	ttactgtcac	gagcggccca	taatcgcgct	5760
ttttttttaa	aaggcgcgag	acagcaaaca	ggaagctcgg	gtttcaacct	tcggagtggt	5820
cgcatgctg	gagactggat	ctttacaata	cagtaaggca	agccaccatc	tgcttcttag	5880
gtgcatgcca	cggatccac	gtgcagaaca	acatagtctg	aagaaggggg	ggaggagcat	5940
gttcattctc	tgtagcagta	agagcttggt	gataatgacc	aaaactggag	tctcgaaatc	6000
atataaatag	acaatatatt	ttcacacaat	gagatttgta	gtacagtctc	attctctctc	6060
ttgcataaat	aagaaattca	tcaagaactt	ggtttgatat	ttcaccaaca	cacacaaaaa	6120
acagtacttc	actaaattta	cacacaaaaa	aaaatggaat	tgaacaacgt	tatcttgga	6180
aaggaaggta	aggttgccgt	tgccaccatc	aacagaccaa	aggttttgaa	tgctttgaac	6240
tctgacactt	tgaagaaaat	ggactacgtc	attggtgaaa	ttgaaaacga	ttctgaagtt	6300
ttggctgtca	tcttgaccgg	tgccggtgaa	aagtctttcg	ttgctggtgc	tgatatctct	6360
gaaatgaagg	aaatgaacac	cattgaaggt	agaaagttcg	gtatcttagg	taacaaggtt	6420
ttcagaagat	tggaattggt	ggaaaagcca	gtcattgtctg	ctgtcaacgg	tttcgctttg	6480
gggtggtggt	gtgaaattgc	catgtcctgt	gacatcagaa	ttgcttcttc	taacgctcgt	6540
ttcggccaac	cagaagtcgg	tctaggtatc	actccaggtt	tcggtggtac	tcaaagatta	6600
tcagattgg	ttggtatggg	tatggccaag	caattgatct	tcaccgctca	aaacatcaag	6660
gctgacgaag	ctttgagaat	tggttttagtc	aacaaggttg	ttgaaccatc	tgaattgatg	6720
aacactgcc	aggaaattgc	taacaagatc	gtctccaacg	ctccagttgc	tgtaaaattg	6780
tccaagcaag	ccatcaacag	aggatgcaa	tgtgatatcg	acaccgcttt	ggcctttgaa	6840
tctgaagctt	tcggtgaatg	tttctccact	gaagacaaaa	aggatgctat	gaccgctttc	6900
atcgaaaaga	gaaagattga	aggtttcaag	aacaggtgat	gagcccgggc	gcgaatttct	6960
tatgatattat	gattttttatt	attaataaag	ttataaaaaa	aataagtgtg	tacaaatttt	7020
aaagtgactc	ttaggtttta	aaacgaaaat	tcttattctt	gagtaactct	ttcctgtagg	7080
tcagggtgct	ttctcaggta	tagcatgagg	tcgctcttat	tgaccacacc	tctaccggca	7140
tgccgagcaa	atgcctgcaa	atcgctcccc	atttcaccca	attgtagata	tgctaaactc	7200
agcaatgagt	tgatgaatct	cgggtgtgat	tttatgtcct	cagaggacaa	cacctgttgt	7260
aatcgttctt	ccacacggat	ccacagccta	gccttcagtt	gggctctatc	ttcatcgtca	7320
ttcattgcat	ctactagccc	cttacctgag	cttcaagacg	ttatatcgct	tttatgtatc	7380
atgatcttat	cttgagatat	gaatacataa	atatatttac	tcaagtgtat	acgtgcatgc	7440
tttttttacg	gtttaaacat	ttaaatgggc	cgctctagag	gatccccggg	taccgagctc	7500

-continued

gggccagcg	ctactagttc	cggtaatgtg	aaaacaaacc	cggtctcgaa	gcggagatcc	7560
ggcgataatt	accgcagaaa	taaaccata	cacgagacgt	agaaccagcc	gcacatggcc	7620
gggaaaaactc	ctgcgagaat	ttcgtaaact	cgcgcgcat	gcacatgtat	ttcctaagtc	7680
ggcacttcca	ggcctcgaga	cctctgacat	gcttttgaca	ggaatagaca	ttttcagaat	7740
gttatccata	tgcccttcgg	gtttttttcc	ttccttttcc	atcatgaaaa	atctctcgag	7800
accgtttatc	cattgctttt	ttgttgcttt	tttccctcgt	tcacagaaag	tctgaagaag	7860
ctatagtaga	actatgagct	ttttttgttt	ctgttttcc	tttttttttt	tttacctctg	7920
tggaattgt	tactctcaca	ctctttagtt	cgtttggttg	ttttgtttat	tccaattatg	7980
accggtgacg	aaacgtgggc	gatggtgggt	accgcttatg	ctccctccca	ttagtttcga	8040
ttatataaaa	aggccaaata	ttgtattatt	ttcaaagtc	ctatcattat	cgtctaact	8100
ctaatttctc	ttaaattttt	tctctttctt	tcctataaca	ccaatagtg	aaatcttttt	8160
ttctttctata	tctacaaaa	cttttttttt	ctatcaacct	cgttgataaa	ttttttcttt	8220
aacaatcggt	aataattaat	taattggaag	ataaccattt	tttctctctt	ttatacacac	8280
attcaaaaga	aagaaaaaaa	atatacccca	gctagttaaa	gaaaatcatt	gaaaagaata	8340
agaagataag	aaagatttaa	ttatcaaa	atatcaatat	gcctcaatcc	tggaagaac	8400
tgcccgctga	taagcgcgcc	cgccctcgaa	aaaccatccc	tgatgaatgg	aaagtccaga	8460
cgctgctgc	ggaagacagc	gttattgatt	tcocaaagaa	atcggggac	ctttcagagg	8520
ccgaactgaa	gatcacagag	gcctccgctg	cagatcttgt	gtccaagctg	gcggccggag	8580
agtgacctc	ggtggaagtt	acgctagcat	tctgtaaacg	ggcagcaatc	gcccagcagt	8640
taacaaactg	cgccacagag	ttcttccctg	acgcccgtct	cgcgaggca	agggaactcg	8700
atgaatacta	cgaaagcac	aagagaccg	ttggtccact	ccatggcctc	cccatctctc	8760
tcaagacca	gcttcgagtc	aagggtacg	aaacatcaat	gggctacatc	tcatggctaa	8820
acaagtacga	cgaaggggac	tcggttctga	caaccatgct	ccgcaaagcc	ggtgcccgtct	8880
tctacgtcaa	gacctctgtc	ccgcagaccc	tgatggctctg	cgagacagtc	aacaacatca	8940
tcgggcgcac	cgtaaccca	cgcaacaaga	actggctctg	cgggcgagtc	tctggtggtg	9000
agggtgcgat	cggtgggatt	cggtggcg	tcacgggtgt	aggaaaggat	atcggtggct	9060
cgattcgagt	gcggcccg	ttcaacttcc	tgtaggtct	aaggccgagt	catggcgggc	9120
tgccgtatgc	aaagatggcg	aacagcatgg	agggtcagga	gacggtgcac	agcgttgctg	9180
ggccgattac	gcactctgtt	gaggacctcc	gcctcttcac	caaaccgctc	ctcggtcagg	9240
agccatggaa	atacgactcc	aaggctcatcc	ccatgcctctg	gcgccagtc	gagtcggaca	9300
ttattgcctc	caagatcaag	aacggcgggc	tcaatatcgg	ctactacaac	ttcgacggca	9360
atgtccttcc	acacctctct	atcctgcgcg	gcgtggaaac	caccgtcgcc	gcactcgcca	9420
aagccggtca	caccgtgacc	ccgtggagcg	catacaagca	cgatttcggc	cacgatctca	9480
tctcccatat	ctacgcggct	gacggcagcg	ccgacgtaat	gcgcgatatc	agtgcacccg	9540
gcgagccggc	gattccaaat	atcaaagacc	tactgaaccc	gaacatcaaa	gctgttaaca	9600
tgaacgagct	ctgggacagc	catctccaga	agtgaatta	ccagatggag	taccttgaga	9660
aatggcgggg	ggctgaagaa	aaggccggga	aggaaactgga	cgccatcatc	gcgccgatta	9720
cgccatccgc	tgccgtacgg	catgaccagt	tcgggtacta	tggtgatgcc	tctgtgatca	9780

-continued

acctgctgga	tttcacgagc	gtggttggtc	cggttacett	tgcggataag	aacatcgata	9840
agaagaatga	gagtttcaag	gcggttagtg	agcttgatgc	cctcgtgcag	gaagagtatg	9900
atccggaggc	gtaccatggg	gcaccggttg	cagtgcaggt	tatcggacgg	agactcagtg	9960
aagagaggac	gttggcgatt	gcagaggaag	tggggaagtt	gctgggaaat	gtggtgactc	10020
cataggtcga	gaatttatac	ttagataagt	atgtacttac	aggtatatatt	ctatgagata	10080
ctgatgtata	catgcatgat	aatattttaa	cggttattag	tgccgattgt	cttgtgcgat	10140
aatgacgttc	ctatcaaagc	aatacactta	ccacctatta	catgggcca	gaaaatattt	10200
tcgaacttgt	ttagaatatt	agcacagagt	atatgatgat	atccgttaga	ttatgcatga	10260
ttcattccta	caacttttcc	gtagcataag	gattaattac	ttggatgcca	ataaaaaaaaa	10320
aaaacatcga	gaaaatttca	gcctgctcag	aaacaattgc	agtgtatcaa	agtaaaaaaaaa	10380
agattttcgc	tacatgttcc	ttttgaagaa	agaaaatcat	ggaacattag	atttacaaaa	10440
atttaaccac	cgctgattaa	cgattagacc	gttaagcgca	caacaggtta	ttagtacaga	10500
gaaagcattc	tgtggtgttg	ccccggactt	tcttttgcca	cataggtaaa	tcgaatacca	10560
tcatactatc	ttttccaatg	actccctaaa	gaaagactct	tcttcgatgt	tgtatacggt	10620
ggagcatagg	gcaagaattg	tggcttgaga	tgaattcact	ggccgtcggt	ttacaacgtc	10680
gtgactggga	aaaccttgcc	gttaccaca	ttaatcgctt	tgcagcacat	ccccctttcg	10740
ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	10800
tgaatggcga	atggcgccgt	atgcggtatt	ttctccttac	gcactctgtc	ggtattttcac	10860
accgcatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	agccagcccc	10920
gacacccgcc	aacacccgct	gacgcgcctt	gacgggcttg	tctgctcccg	gcacccgctt	10980
acagacaagc	tgtgaccgtc	tccgggagct	gcactgtgca	gaggttttca	ccgtcatcac	11040
cgaacgcgcg	gagacgaaag	ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	11100
taataatggt	ttcttagacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	11160
tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	11220
aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	11280
ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggtga	11340
aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggt	ttacatcgaa	ctggatctca	11400
acagcggtaa	gaccccttag	agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	11460
ttaaagtctt	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	gagcaactcg	11520
gtcgccgcac	acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	11580
atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgcataaacc	atgagtgata	11640
acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	11700
tgcacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	11760
ccataccaaa	cgacgagcgt	gacaccacga	tgcctgtagc	aatggcaaca	acgttgcgca	11820
aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	11880
aggcggataa	agttgcagga	ccacttctgc	gctcgccctt	tccggctggc	tggtttattg	11940
ctgataaatc	tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	12000
atggtaagcc	ctcccgatc	gtagtattct	acacgacggg	gagtcaggca	actatggatg	12060

-continued

```

aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag 12120
accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga 12180
tctaggtgaa gatccttttt gataatctca tgacccaaat cccttaacgt gagttttcgt 12240
tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat cctttttttc 12300
tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc 12360
cggatcaaga gctaccaact ctttttcgga aggtaactgg cttcagcaga gcgcagatac 12420
caaatactgt ctttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac 12480
cgctacata cctcgcctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt 12540
cgtgtcttac cgggttggaac tcaagacgat agttaccgga taaggcgag cggtcgggct 12600
gaacggggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat 12660
acctacagcg tgagctatga gaaagcgcca cgcttccgga agggagaaaag gcggacaggt 12720
atccggtaa cggcagggtc ggaacaggag agcgacgag ggagcttcca gggggaaacg 12780
cctggtatct ttatagtct gtcgggttcc gccacctctg acttgagcgt cgatttttgt 12840
gatgctcgtc agggggggcg agcctatgga aaaacgccag caacgcggcc tttttacggt 12900
tcttgccctt ttgctggcct tttgtcaca tgttctttcc tgcttatcc cctgattctg 12960
tggataaccg tattaccgcc tttgagttag ctgataccgc tcgccgcagc cgaacgaccg 13020
agcgagcgga gtcagtgagc gaggaagcgg aagagcgccc aatacgcaaa ccgcctctcc 13080
ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac tggaaagcgg 13140
gcagtgagcg caacgcgaatt aatgtgagtt agctcactca ttaggcaccc caggctttac 13200
actttatgct tccggctcgt atgttggttg gaattgtgag cggataacaa tttcacacag 13260
gaaacagcta tgaccatgat tacgcc 13286

```

```

<210> SEQ ID NO 42
<211> LENGTH: 16359
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: pBOL36

```

```

<400> SEQUENCE: 42

```

```

aagcttgcat gcctgcaggc cgacggcgcg ccgggcccgt ttaacaatg gcaaactgag 60
cacaacaata ccagtcggga tcaactggca ccactctctc cgtagtctca tctaattttt 120
cttcgggatg aggttcgaga tatacggcaa cacctttatt atggtttccc tgagggaata 180
atagaatgtc ccattcgaaa tcaccaattc taaacctggg cgaattgtat ttcgggtttg 240
ttaactcgtt ccagtcagga atgttcacg tgaagctatc ttccagcaaa gtctccactt 300
cttcacaaa ttgtgggaga atactcccaa tgctcttata tatgggactt ccgggaaaca 360
cagtaccgat acttcccaat tcgtcttcag agtcattgt ttgtttgaag agactaatca 420
aagaatcgtt ttctcaaaaa aattaatatc ttaactgata gtttgatcaa agggggcaaaa 480
cgtaggggca aacaaacgga aaaatcggtt ctcaaatttt ctgatgcaa gaactctaac 540
cagtcttata taaaaattgc cttatgatcc gtctctccgg ttacagcctg tgtaactgat 600
taatcctgcc tttctaatac ccattctaata gttttaatta agggattttg tcttcattaa 660
cggctttcgc tcataaaaaa gttatgacgt tttgcccgca ggcgggaaac catccacttc 720

```

-continued

acgagactga	tctcctctgc	cggaacaccg	ggcatctcca	acttataagt	tggagaaata	780
agagaatttc	agattgagag	aatgaaaaaa	aaaaaaaaaa	aaaaggcaga	ggagagcata	840
gaaatggggt	tacttttttg	gtaaagctat	agcatgccta	tcacatataa	atagagtgcc	900
agtagcgact	tttttcacac	tcgaaatact	cttactactg	ctctcttggt	gtttttatca	960
cttcttggtt	cttcttggtg	aatagaatat	caagctacaa	aaagcataca	atcaactatc	1020
aactattaac	tatatcgtaa	tacacaggcc	ggccaaaatg	aaggccaaat	caaggcggga	1080
agggacaacc	aggacgtaaa	gggtagcctc	cccataacat	aaactcaata	aaatatatag	1140
tcttcaactt	gaaaaaggaa	caagctcatg	caaagagggtg	gtacccgcac	gccgaaatgc	1200
atgcaagtaa	cctattcaaa	gtaatatctc	atacatgttt	catgagggtg	acaacatgcg	1260
actgggtgag	catatgttcc	gctgatgtga	tgtgcaagat	aaacaagcaa	gacagaaaact	1320
aacttcttct	tcatgtaata	aacacacccc	gcgtttatct	acctatcttt	aaacttcaac	1380
accttatatc	ataactaata	tttcttgaga	taagcacact	gcaccatac	cttccttaaa	1440
aacgtagctt	ccagtttttg	gtggttctgg	cttccttccc	gattccgccc	gctaaacgca	1500
taattttggt	gcctgggtgc	atttgcaaaa	tgcataacct	atgcatttaa	aagattatgt	1560
atgctcttct	gacttttctg	gtgatgaggg	tcgtggaaaa	aatgaataat	ttatgaattt	1620
gagaacaatt	ttgtgttggt	acggtatctt	actatggaat	aatcaatcaa	ttgaggattt	1680
tatgcaaata	tcgtttgaat	atttttccga	ccctttgagt	acttttcttc	ataattgcat	1740
aatattgtcc	gctgcccgtt	tttctggtag	acggtgtctt	gatctacttg	ctatcgttca	1800
acaccacett	attttctaac	tatttttttt	ttagctcatt	tgaatcagct	tatggtgatg	1860
gcacattttt	gcataaacct	agctgtcctc	gttgaacata	ggaaaaaaa	atatataaac	1920
aaggctcttt	cactctcctt	ggaatcagat	ttgggtttgt	tccttttatt	ttcatatttc	1980
ttgtcatatt	cttttctcaa	ttattatctt	ctactcataa	cctcacgcaa	aataacacag	2040
tcaaatcaat	caaaatggac	ttcaacttga	ccagagaaca	agaattgggtc	agacaaatgg	2100
ttagagaatt	tgctgaaaac	gaagttaagc	caattgctgc	tgaatcgat	gaaactgaaa	2160
gattcccaat	ggaaaacgtc	aagaagatgg	gtcaatacgg	tatgatgggt	attccattct	2220
ctaaggaata	cggtgggtgc	ggtggtgacg	tcttgtctta	catcattgct	gtcgaagaat	2280
tgtccaaggt	ttgtggtacc	actggtgtca	tcttatctgc	tcacacttct	ctatgtgcct	2340
ccttgatcaa	cgaacacggg	actgaagaac	aaaagcaaaa	gtacttggtt	ccattggcca	2400
agggtgaaaa	gattggtgcc	tacggtttga	ctgaaccaa	cgctgggtact	gactctggtg	2460
ctcaacaaac	tgttgccgtt	ttggaagggtg	accactacgt	catcaacggg	tccaagatct	2520
tcatacccaa	cggtgggtgt	gctgacacct	ttgtcatctt	cgctatgacc	gacgtacca	2580
agggtagcaa	gggtatctct	gctttcatta	ttgaaaaggg	tttcaagggt	ttctccatcg	2640
gtaagggtcg	acaaaagttg	ggtatcagag	cttcctctac	cactgaattg	gttttcgaag	2700
acatgattgt	tccagttgaa	aacatgatcg	gtaagggaagg	taagggtttc	ccaattgcca	2760
tgaagacttt	agatgggtgt	agaattggta	ttgctgctca	agctttgggt	attgctgaag	2820
gtgccttcaa	cgaagctaga	gcttacatga	aggaaagaaa	gcaattcggg	agatctttgg	2880
acaaattcca	aggtttggct	tggatgatgg	ctgacatgga	cggtgccatc	gaatctgctc	2940
gttacttggt	ctacaaggct	gcttacttga	agcaagctgg	tttgccatac	accgtcgatg	3000

-continued

ctgccagagc taagttgcac gctgccaacg ttgccatgga tgtcaccacc aaggtgtcc	3060
aattattcgg tggttacggt tacaccaagg actaccagc tgaagaatg atgagagatg	3120
ctaagatcac tgaatctac gaaggtactt ctgaagttca aaagttgggt atctccggtg	3180
agatcttcag ataggcccg gcataaagca atcttgatga ggataatgat tttttttga	3240
atatacataa atactaccgt ttttctgcta gattttgtga agacgtaaat aagtacatat	3300
tactttttaa gccaagacaa gattaagcat taactttacc cttttctctt ctaagtttca	3360
atactagtta tcaactgtta aaagttatgg cgagaacgtc ggcgggttaa atattattacc	3420
ctgaacgtgg tgaattgaag ttctaggatg gtttaagat ttttcctttt tgggaaataa	3480
gtaacaata tattgctgcc ttgcaaac gcacataccc acaatatgtg actattggca	3540
aagaacgcac tatcctttga agaggtggat actgatacta agagagctct tattccggct	3600
ccacttttag tccagagatt acttgctctc ttacgtatca gaacaagaaa gcatttccaa	3660
agtaattgca ttgcccctg agcagtatat atatactaag aagtttaaac atttaaacgg	3720
ccggcctaga aagcatacta tactattcga cacttccttt caatcctgga attaacagtc	3780
acttttaaaa aagacatcta ccgtgaaggt gccgtagagt atcgcgttac catatcgcca	3840
aaaaactgata tacgcccgcg aaaccaggca aacaattgaa aagaaaaatt ttgaggaaact	3900
ctctgcacg aagccgtcta gagttaccac tagtcagatg ccgcgggcac ttgagcacct	3960
catgcacagc aataacacaa cacaatggtt agtagcaacc tgaattcggg cattgatgca	4020
tgcattgtcc gtgaagcggg acaaccagaa aagtcgtcta taaatgccg cagtgcgat	4080
cacgtggcg gggttttaag agtgcatac acaaatgtc gcattaccgc ggaaccgcca	4140
gatattcatt acttgacgca aaagcgtttg aaataatgac gaaaaagaag gaagaaaaa	4200
aaagaaaaat accgcttcta ggcgggttat ctactgatcc gagcttcac taggatagca	4260
cccaaacacc tgcataattg gacgacctt acttacacca ccaaaaacca ctttcgcctc	4320
tccgcacct gataacgtcc actaatgag cgattacctg agcggctctc tttgtttgc	4380
agcatgagac ttgcatactg caaatcgtaa gtagcaacgt ctcaaggtca aaactgtatg	4440
gaaacctgt cactcactt aattctagct agcctaccct gcaagtcaag aggtctccgt	4500
gattcctagc cactcaagg tatgcctctc ccgggaaact gtggcctttt ctggcacaca	4560
tgatctccac gatttcaaca tataaatagc ttttgataat ggcaatatta atcaaattha	4620
ttttacttct ttctgttaac atctctcttg taatccctta ttccttctag ctatttttca	4680
taaaaaacca agcaactgct tatcaacaca caaacactaa atcaaatgg tcgatttcga	4740
atactctatc ccaaccagaa tcttcttcgg taaggacaag atcaacgttt tgggtagaga	4800
attgaagaaa tacgggtcca aggttttgat tgtctacggt ggtggttcca tcaagagaaa	4860
cgggtatctac gacaaggctg tctccatttt ggaaaagaac tctatcaaat tctacgaatt	4920
ggctggtgtt gaaccaaacc caagagttac caccgtcgaa aagggtgtca agatctgtcg	4980
tgaaacgggt gttgaagttg ttttgcccat cgggtggtgt tctgccattg actgtgccaa	5040
ggcattgtct gctgcctgtg aatacgatgg taacccatgg gacattgtct tggatggttc	5100
taagatcaag cgtgtcttac caattgcttc catcttgact atcgtctgta ctggttctga	5160
aatggacacc tgggctgtta tcaacaacat ggacactaac gaaaagtga ttgctgtcga	5220
cccagatatg gcccacaaat tctctattht ggacccaacc tacacttaca ctgttccaac	5280

-continued

caaccaaact gctgctggta ctgctgatat catgtctcac atctttgaag tttacttctc	5340
caacaccaag accgcttact tgcaagacag aatggctgaa gctctattaa gaacctgtat	5400
caagtagcgt ggtattgctt tggaaaagcc agatgactac gaagccagag ctaacttgat	5460
gtgggcttcc tctttggcta tcaacgggtt attgacttac ggtaaggaca ccaactggtc	5520
cgttcatttg atggaacacg aattgtctgc ttactacgat atcactcacg gtgtcgggtt	5580
ggccatcttg actccaaact ggatggaata cattttgaac aacgacactg tctacaagtt	5640
cgtcgaatac ggtgttaacg tctgggggat tgacaaggaa aagaaccact acgacattgc	5700
tcaccaagcc atccaaaaga ccagagacta tttcgtcaac gttttgggtt taccatccag	5760
attaagagat gttggtattg aagaagaaaa attggatata atggctaagg aatctgtcaa	5820
attgactggt ggtaccattg gtaacttgag acctgttaac gcttctgaag ttttgcaa	5880
cttcaagaaa tctgtttagg cccgggctcc tgttgaagta gcatttaac ataatttttg	5940
tcacatttta atcaacttga tttttctggt ttaatttttc taattttaat ttttaatttt	6000
ttatcaatgg gaactgatac actaaaaaga attaggagcc aacaagaata agccgcttat	6060
ttctactag agtttgctta aaatttcac tcgaattgac attctaata tttatccaca	6120
cacacacctt aaaattttta gattaaatgg catcaactct tagcttcaca cacacacaca	6180
caccgaagct ggttggttta tttgatttga tataattggt ttctctggat ggtacttttt	6240
ctttcttggt tatttcttat tttaaaat gaaacgcaca caagtcataa ttattcta	6300
agagcacaat tcacaacacg cacatttcaa ctttaatat tttttagaaa cactttattt	6360
agtcataatc ttaattttta atatatataa tgcacacaca cgtttaaatg ggcccgccgc	6420
ccgtttaaac ggccggccct tcccttttac agtgcttcgg aaaagcacag cgttgtccaa	6480
gggaacaatt tttcttcaag ttaatgcata agaaatatct ttttttatgt ttagctaagt	6540
aaaagcagct tggagtaaaa aaaaaaatga gtaaatctct cgatggatta gtttctcaca	6600
ggtaacatag caaaaaccaa gaaaagcccg cttctgaaaa ctacagtga cttgtatgct	6660
aaagggccag actaatggga ggagaaaaag aaacgaatgt atatgctcat ttacactcta	6720
tatcaccata tggaggataa gttgggctga gcttctgac caatttatc tatccattag	6780
ttgtgatata gtccaccag ccaacacttg atagtatcta ctgccattc acttccagca	6840
gcgccagtag ggttggttag cttagtataa atgtgcgcac cacaagccta catgactcca	6900
cgtcacatga aaccacaccc tggggccttg ttgcgctagg aataggatat gcgacgaaga	6960
cgttcttgct tagtaaccac accacatttt cagggggtcg atctgcttgc ttcctttact	7020
gtcacgagcg gcccataatc gcgctttttt tttaaaaggc gcgagacagc aaacaggaag	7080
ctcgggtttc aaccttcgga gtggtcgcag atctggagac tggatcttta caatacagta	7140
aggcaagcca ccatctgctt cttaggtgca tgcgacggta tccacgtgca gaacaacata	7200
gtctgaagaa gggggggagg agcatgttca ttctctgtag cagtgaagac ttggtgataa	7260
tgacaaaaac tggagtctcg aaatcatata aatagacaat atattttcac acaatgagat	7320
ttgtagtaca gttctattct ctctcttgca taaataagaa attcatcaag aacttggttt	7380
gatattttcac caacacacac aaaaaacagt acttacttaa atttacacac aaaacaaaat	7440
gaaggttacc aacaaaaagg aattgaagca aaagttgaac gaattgagag aagctcaaaa	7500
gaagttcgct acctacactc aagaacaagt tgacaagatc ttcaagcaat gtgccattgc	7560

-continued

tgctgccaag gaacgtatca acttggccaa gttggctgtc gaagaaaccg gtattggttt	7620
ggttgaagac aagatcatca agaaccactt cgctgctgaa tacatctaca acaagtacaa	7680
gaacgaaaag acctgtggta tcatcgacca cgatgactct ttgggtatca ccaaggttgc	7740
tgaaccaatc ggtattgtcg ccgccattgt cccaaccact aacccaactt ccactgccat	7800
cttcaaactt ttgatctcct tgaagaccag aaacgctatc ttcttctccc cacaccaag	7860
agccaagaag tccaccattg ctgctgccaa attaactctg gatgctgctg ttaaggctgg	7920
tgccccaaag aacattattg gttggatcga tgaaccttcc attgaattgt ctcaagactt	7980
gatgtctgaa gctgatatca tcttggttac cgggtgtcca tccatggtca aggccgtta	8040
ctcttctggt aagccagcta ttggtgttgg tgctggtaac actccagcta tcatcgatga	8100
atctgctgac attgacatgg ctgtctcttc cattatcttg tccaagactt atgacaacgg	8160
tgcatctgt gctctgaac aatccatctt gggtatgaac tctatctacg aaaaggtaa	8220
ggaagaattt gttaagagag gttcctacat cttaaaccaa aatgaaattg ccaagatcaa	8280
ggaaccatg ttcaagaacg gtgccatcaa cgctgacatt gtcggtaaat ctgcttacct	8340
cattgccaa atggctggta ttgaagtcc acaaccact aagattttga tcggtgaagt	8400
tcaatctgtc gaaaagtctg aattattctc tcacgaaaag ttgtctccag tcttggtat	8460
gtacaaggtc aaggatttgc acgaagcttt gaagaaggct caaagattaa ttgaattagg	8520
tggttctggt cacacctctt ctctatacat tgactctcaa aacaacaagg acaaggtaa	8580
ggaattcggc ctagtctatga agacttccag aactttcatc aacatgccat cttctcaagg	8640
tgcttctggt gatttgtaca actttgccat tgctccatct ttcactttag gttgtggtac	8700
ctgggtggt aactctgttt ctcaaaacgt tgaaccaaag catttgctaa acatcaagtc	8760
cgttgctgaa agaagagaaa acatgtttgt gttcaagggt ccacaaaaga tctacttcaa	8820
atacggttgt ttgagatttg ctttgaagga attgaaagat atgaacaaga agcgtgcttt	8880
catcggtact gacaaggatt tgttcaaatt gggttacgtt aacaagatca ctaaggtttt	8940
ggatgaaatt gatatacagt actccatctt cactgatatc aaatctgacc caaccattga	9000
ctccgtcaag aagggtgcta aggaaatgtt gaacttcgaa ccagatacca ttatctccat	9060
tggtggtggt tctccaatgg atgctgccaa gggtatgcat ttgtgtacg aataccaga	9120
agctgaaatc gaaaacttgg ccatcaactt catggacatc agaagagaa tctgtaactt	9180
cccaaagttg ggtaccaagg ccatttctgt tgccattcca accaccgtg gtaccggttc	9240
tgaagctact ccatttctgt tcatcaccaa cgacgaaacc ggatgaagt acccattgac	9300
ctcttacgaa ttgactccaa acatggccat cattgacact gaattgatgt tgaacatgcc	9360
aagaagttg actgctgcta ccggtattga cgcttttagc cacgctatcg aagcttacgt	9420
ctccgttatg gccactgact aactgacga attggctttg agagctatca agatgatctt	9480
caagtacttg ccaagagctt acaagaacgg tactaacgat atcgaagctc gtgaaaagat	9540
ggctcacgct tccaacattg ctggtatggc ttctgctaac gctttcttgg gtgtttgtca	9600
ctccatggcc cacaagttgg gtgctatgca ccacgttctc cacggtattg cttgtgctgt	9660
tttgattgaa gaagtcatca agtacaacgc tactgactgt ccaaccaagc aaactgcttt	9720
cccacaatac aagtctccaa acgccaagag aaagtacgct gaaattgctg aatacttgaa	9780
cttgaaaggt acttctgaca ctgaaaaggt cactgcttta atcgaagcta tctccaagtt	9840

-continued

gaagattgac	ttatctatct	ctcaaaacat	ctctgctgct	ggatattaaca	agaaggactt	9900
ctacaacact	ttagacaaga	tgtccgaatt	ggctttcgat	gaccaatgta	ccaccgctaa	9960
cccaagatac	ccattgatct	ctgaattgaa	ggatatctac	atcaagtcct	tttaagcccc	10020
ggcgcggtac	tcttatgtct	ttacgattta	tagttttcat	tatcaagtat	gcctatatata	10080
gtatatagca	tcttttagatg	acagtgttcg	aagtttcacg	aataaaagat	aatattctac	10140
tttttgctcc	caccgcgttt	gctagcacga	gtgaacacca	tcctctgcct	gtgagttgta	10200
cccatctctc	taaaactgtag	acatggtagc	ttcagcagtg	ttcggttatgt	acggcatcct	10260
ccaacaaaca	gtcgggtata	gtttgtcctg	ctcctctgaa	tcgtctccct	cgatatttct	10320
cattttccct	cgcatgccag	cattgaaatg	atcgaagttc	aatgatgaaa	cggtaattct	10380
tctgtcatct	actcatctca	tctcatcaag	ttatataatt	ctatacggat	gtaatttttc	10440
acttttcgct	ttgacgtcca	ccctataatt	tcaattattg	aacctcaca	aatgatgcac	10500
tgcaatgtac	acacctcat	atagttttaa	catttaaattg	ggcgctcta	gaggatcccc	10560
gggtaccgag	ctcgggcccc	gcgctactag	ttccggtaat	ttgaaaacaa	acccggctct	10620
gaagcggaga	tccggcgata	attaccgcag	aaataaaccc	atacacgaga	cgtagaacca	10680
gccgcacatg	gccggagaaa	ctcctgcgag	aatttcgtaa	actcgcgcgc	attgcatctg	10740
tatttctcaa	tgcggcactt	ccaggcctcg	agacctctga	catgcttttg	acaggaatag	10800
acattttcag	aatgttatcc	atatgccttt	cgggtttttt	tccttccttt	tccatcatga	10860
aaaaatctct	gagaccgttt	atccattgct	tttttggtgt	ctttttccct	cgttcacaga	10920
aagtctgaag	aagctatagt	agaactatga	gctttttttg	tttctgtttt	cctttttttt	10980
ttttttacct	ctgtggaaat	tgttactctc	acactcttta	gttcgtttgt	ttgttttggt	11040
tattccaatt	atgaccggtg	acgaaacgtg	gtcgatggtg	ggtagccgctt	atgctccctt	11100
ccattagttt	cgattatata	aaaaggccaa	atattgtatt	attttcaa	atgtctatcat	11160
tatcgtctaa	catctaattt	ctcttaaat	ttttctcttt	ctttcctata	acaccaatag	11220
tgaaaatctt	tttttctct	atatctacaa	aaactttttt	tttctatcaa	cctcgttgat	11280
aaattttttc	tttaacaate	gttaataatt	aattaattgg	aaaataacca	ttttttctct	11340
cttttataca	cacattcaaa	agaaagaaaa	aaaatatacc	ccagctagtt	aaagaaaatc	11400
attgaaaaga	ataagaagat	aagaaagatt	taattatcaa	acaatatcaa	tatgcctcaa	11460
tcttggaag	aactggccgc	tgataagcgc	gccgcctcg	caaaaacat	cctgatgaa	11520
tgaaaagtcc	agacgctgcc	tgcggaagac	agcgttattg	atttccaaa	gaaatcgggg	11580
atcctttcag	aggccgaact	gaagatcaca	gaggcctcgc	ctgcagatct	tgtgtccaag	11640
ctggcgccgc	gagagttgac	ctcggtggaa	gttacgctag	cattctgtaa	acgggcagca	11700
atcgccagc	agttaacaaa	ctgcgccac	gagttcttcc	ctgacgcgc	tctcgcgag	11760
gcaagggaac	tcatgaata	ctacgcaaag	cacaagagac	ccgttggtcc	actccatggc	11820
ctccccatct	ctctcaaaga	ccagcttcga	gtcaagggtc	acgaaacatc	aatgggctac	11880
atctcatggc	taaacaagta	cgacgaagg	gactcggttc	tgacaacat	gctccgcaaa	11940
gccgggtgcc	tctttctacgt	caagacctct	gtccgcgaga	ccctgatggg	ctgcgagaca	12000
gtcaacaaca	tcatcgggcg	caccgtcaac	ccacgcaaca	agaactggtc	gtcgcgccgc	12060
agttctgggtg	gtgaggggtgc	gatcggtggg	attcgtgggtg	gcgtcatcgg	tgtagggaacg	12120

-continued

gatatcgggtg gctcgattcg agtgccggcc gcgttcaact tcctgtacgg tctaaggccg	12180
agtcatgggc ggctgccgta tgcaaatgatg gcgaacagca tggaggggtca ggagacgggtg	12240
cacagcggtt tggggccgat tacgcactct gttgaggacc tccgcctctt caccaaatcc	12300
gtcctcggtc aggagccatg gaaatacgac tccaagggtca tccccatgcc ctggcgccag	12360
tccgagtcgg acattattgc ctccaagatc aagaacggcg ggctcaatat cggtacttac	12420
aacttcgacg gcaatgtcct tccacaccct cctatcctgc gcggcggtga aaccaccgtc	12480
gccgcactcg ccaaagccgg tcacaccgtg accccgtgga cgccatacaa gcacgatttc	12540
ggccacgatc tcatctccca tatctacgcg gctgacggca gcgccgacgt aatgcgcgat	12600
atcagtgcac ccggcgagcc ggcgattcca aatatcaaag acctactgaa ccggaacatc	12660
aaagctgtta acatgaacga gctctgggac acgcatctcc agaagtggaa ttaccagatg	12720
gagtaccttg agaaatggcg ggaggtctga gaaaaggccg ggaaggaaact ggacgccatc	12780
atcgcgccga ttacgcctac cgctgcggta cggcatgacc agttccggta ctatgggtat	12840
gcctctgtga tcaacctgct ggatttcacg agcgtggttg ttccggttac ctttgcggtat	12900
aagaacatcg ataagaagaa tgagagtctc aaggcgggtta gtgagcttga tgccctcgtg	12960
cagggaagagt atgatccgga ggcgtaacct ggggcaccgg ttgcagtga ggttatcggga	13020
cggagactca gtgaagagag gacgttggcg attgcagagg aagtggggaa gttgctggga	13080
aatgtggtga ctccataggt cgagaattta tacttagata agtatgtact tacaggtata	13140
tttctatgag atactgatgt atacatgcat gataatattt aaacggttat tagtgccgat	13200
tgtcttgtgc gataatgacg ttcctatcaa agcaatacac ttaccaccta ttacatgggc	13260
caagaaaata ttttcgaact tgtttagaat attagcacag agtatatgat gatatccgtt	13320
agattatgca tgattcatc ctacaacttt ttcgtagcat aaggattaat tacttggatg	13380
ccaataaaaa aaaaaaacat cgagaaaatt tcagcatgct cagaacaat tgcagtgtat	13440
caaagtaaaa aaaagatttt cgctacatgt tccttttgaa gaaagaaaat catggaacat	13500
tagatttaca aaaatttaac caccgctgat taacgattag accgtaagc gcacaacagg	13560
ttattagtag agagaaagca ttctgtggtg ttgccccgga ctttcttttg cgacataggt	13620
aaatcgaata ccatcact atcttttcca atgactccct aaagaaagac tcttcttcga	13680
tgttgatata gttggagcat agggcaagaa ttgtggcttg agatgaattc actggccgtc	13740
gttttacaac gtcgtgactg ggaaaaccct ggcgttacc aacttaatcg ccttcagca	13800
catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa	13860
cagttgcgca gcctgaatgg cgaatggcg ctagtgcggt attttctcct tacgcactcg	13920
tgcggtattt cacaccgcat atggtgcact ctgagtacaa tctgctctga tgccgatag	13980
ttaagccagc cccgacacc gccaacacc gctgacgcgc cctgacgggc ttgtctgctc	14040
ccggcatccg cttacagaca agctgtgacc gtctccggga gctgcattgt tcagaggttt	14100
tcacgcgtcat caccgaaacg cgcgagacga aagggcctcg tgatacgctt atttttatag	14160
gttaatgtca tgataataat ggtttcttag acgtcagggt gcacttttcg gggaaatgtg	14220
cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc gtcgatgaga	14280
caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat	14340
ttcgtgtcgc cccttattcc cttttttgcg gcattttgcc ttctgtttt tgctcaccga	14400

-continued

```

gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc 14460
gaactggatc tcaacagcgg taagatcctt gagagttttc gccccgaaga acgtttttcca 14520
atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgat tgacgccggg 14580
caagagcaac tcggtcgcgc catacactat tctcagaatg acttggttga gtactcacca 14640
gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata 14700
accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag 14760
ctaaccgctt ttttgacaaa catgggggat catgtaactc gccttgatcg ttgggaaccg 14820
gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca 14880
acaacgttgc gaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta 14940
atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct 15000
ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcggg tatcattgca 15060
gcactggggc cagatggtaa gccctcccg atcgtagtta tctacacgac ggggagtcag 15120
gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat 15180
tggttaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt 15240
taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa 15300
cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 15360
gatccttttt ttctgcgcgt aatctgctgc ttgcaacaa aaaaaccacc gctaccagcg 15420
gtggtttgtt tgccggatca agagctacca actcttttcc cgaaggtaac tggcttcagc 15480
agagcgcaga taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag 15540
aactctgtag caccgcctac atacctcgt ctgctaatac tgttaccagt ggtgctgccc 15600
agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 15660
cagcggtcgg gctgaacggg ggggtcgtgc acacagccca gcttgagcg aacgacctac 15720
accgaactga gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga 15780
aaggcggaac ggtatccggt aagcggcagg gtcggaacag gagagcgac gagggagcct 15840
ccagggggaa acgcttggt tctttatagt cctgtcgggt ttcgccacct ctgacttgag 15900
cgtcgatttt tgtgatgctc gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg 15960
gcctttttac ggttcttgcc cttttgctgg ccttttctc acatgttctt tctgcgtta 16020
tccctgatt ctgtggataa ccgtattacc gcctttgagt gagctgatac cgctcgccgc 16080
agccgaacga ccgagcgcag cgagtcagtg agcgaggaag cggaagagcg cccaatacgc 16140
aaacgcctc tccccgcgcg ttggcogatt cattaatgca gctggcacga caggtttccc 16200
gactggaag cgggcagtga gcgcaacgca attaatgtga gttagctcac tcattaggca 16260
ccccaggctt tacactttat gcttcggct cgtatgttgt gtggaattgt gagcggataa 16320
caatttcaca caggaaacag ctatgacat gattacgcc 16359

```

```

<210> SEQ ID NO 43
<211> LENGTH: 8684
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: pBOL113

<400> SEQUENCE: 43

```

-continued

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accataccac agcttttcaa ttcaattcat cttttttttt ttattctttt ttttgatttc	240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg	300
agcacagact tagattggtg tatatacgca tatgtagtgt tgaagaaaca tgaattgcc	360
cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt	420
cgaaagctac atataaggaa cgtgctgcta ctcacctag tcctgttgct gccaaagctat	480
ttaatatcat gcacgaaaag caaacaaact tgtgtgcttc attggatgtt cgtaccacca	540
aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta aaaacacatg	600
tggatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg	660
cgaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca	720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac	780
acggtgtggt gggcccaggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa	840
aggaacctag aggccttttg atgttagcag aattgtcatg caagggtcc ctatctactg	900
gagaatacag taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct	960
ttattgctca aagagacatg ggtggaagag atgaaggtta cgattgggtg attatgacac	1020
ccggtgtggg ttttagtgac aaggagagcg cattgggtca acagtataga accgtggatg	1080
atgtggtctc tacaggatct gacattatta ttgttgaag aggactatct gcaaagggaa	1140
gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctgggaagca tatttgagaa	1200
gatgcggcca gcaaaactaa aaaactgtat tataagtaaa tgcattgata ctaaaactac	1260
aaattagagc ttcaatttaa ttatatcagt tattacccta tgcgggtgta aataccgcac	1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgttaatatt ttgttaaaat	1380
tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa	1440
tcccttataa atcaaaagaa tagaccgaga taggggtgag tgtgttcca gtttgaaca	1500
agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg	1560
gcgatggccc actacgtgaa ccatcaccct aatcaagttt tttgggtcg aggtgccgta	1620
aagcactaaa tcggaacctt aaaggagacc cccgatttag agcttgacgg ggaagccgg	1680
cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa	1740
gtgtagcggg cacgtgcgc gtaaccacca caccgcgcgc gcttaatgcg ccgctacagg	1800
gcgctgcgc ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg	1860
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg	1920
taacgccagg gttttccag tcacgacgtt gtaaaacgac ggccagttag ccgctgtaat	1980
acgactcact ataggcgcaa ttgggtaccg gggccccct cgaggctgac ggtatcgata	2040
agcttgatat cgaattcctg cagcccgggg gatccactag ttctagagcg gccatttaa	2100
acggccggcc ctagatcaga ggggtgtaaa tgaagtgtaa tagtattcat tttcttata	2160
aatcatccct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat	2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat	2280

-continued

cttttaagt	ccaaaaacat	ttgagccgat	ctaaatactt	ctgtgttttc	attaatttat	2340
aaattgtact	cttttaagac	atggaaagta	ccaacatcgg	tgaaacagt	ttttcattta	2400
cttatgggtt	attgggtttt	ccagtgaatg	attatttgtc	gttacccttt	cgtaaaagtt	2460
caaacacggt	tttaagtatt	gtttagtgtc	tctttcgaca	tatatgatta	tcctgcgcg	2520
gctaaagtta	aggatgcaaa	aaacataaga	caactgaagt	taatttacgt	caattaagtt	2580
ttccagggtt	atgatgtttt	gggcttcac	taattcaata	agtatgtcat	gaaatacgtt	2640
gtgaagagca	tccagaaata	atgaaaagaa	acaacgaaac	tgggtcggcc	tgtgttttct	2700
tttctttacc	acgtgatctg	cggcatttac	aggaagtcgc	gcgttttgcg	cagttgttgc	2760
aacgcagcta	cggctaacaa	agcctagtgg	aactcgactg	atgtgttagg	gcctaaaact	2820
gggtgtgaca	gctgaagtga	actattcaat	ccaatcatgt	catggctgtc	acaaagacct	2880
tgccgaccgc	acgtacgaac	acatacgtat	gctaatatgt	gttttgatag	taccagtgga	2940
tcgcagacct	gcaatttttt	tgtaggtttg	gaagaatata	taaaggttgc	actcattcaa	3000
gatagttttt	ttctgtgtgt	tctattcatt	ttattattgt	ttgtttaaat	gttaaaaaaa	3060
ccaagaactt	agtttcaaat	taaattcatc	acacaaacaa	acaaaacaaa	atgaacattg	3120
ttgtttgttt	gaagcaagtt	ccagacactg	ctgaagtcag	aattgacca	gtcaagggtt	3180
ctttaatcag	agaaggtgtt	ccatctatca	tcaaccaga	cgacaagaac	gctttggaag	3240
aagcttttgt	tttgaaggac	aactacggtg	ctcacgttac	cgtcatttcc	atgggtccac	3300
ctcaagccaa	gaacgccttg	gttgaagctt	tggccatggg	tgctgatgaa	gctgtcttat	3360
tgactgacag	agctttcggt	ggtgctgata	cttttagctac	ctctcacacc	attgctgctg	3420
gtatcaagaa	attgaaatac	gatatcgtct	ttgccggtcg	tcaagccatc	gatggtgata	3480
ccgctcaagt	cggctccagaa	attgctgaac	atttgggtat	tccacaagtc	acctacgttg	3540
aaaagggtga	agttgacggt	gacactttga	agatcagaaa	ggcttgggaa	gacgggttacg	3600
aagttgttga	agtcaagact	ccagttctat	tgactgccat	caaggaattg	aacgttccaa	3660
gatacatgtc	cgttgaaaag	atcttcggtg	ctttcgacaa	ggaagtcaag	atgtggactg	3720
ctgatgatat	cgatgtcgac	aaggccaact	tgggtttgaa	aggttctcca	accaaggtca	3780
agaaatcttc	taccaaggaa	gtcaagggtc	aaggtgaagt	cattgacaaa	ccagtcaagg	3840
aagctgccgc	ttacgttgtt	tccaagttga	aggaagaaca	ctacatctaa	agcccgggcg	3900
gagattgata	agacttttct	agttgcata	cttttatatt	taaatcttat	ctattagtta	3960
attttttgta	atttatcctt	atatatagtc	tggttattct	aaaatatcat	ttcagtatct	4020
aaaaattccc	ctcttttttc	agttatatct	taacaggcga	cagtccaaat	gttgatttat	4080
ccagtccega	ttcatcaggg	ttgtgaagca	ttttgtcaat	ggtcgaaatc	acatcagtaa	4140
tagtgccctc	tacttgccct	atagaatttc	tttctcttaa	cgtcaccggt	tggctcttta	4200
tagtttcgaa	atctatgggt	ataccaaag	gtgttcccaa	ttcatcgtaa	cgggcgtatt	4260
ttttaccaat	tgaagtattg	gaatcgtcaa	ttttaagta	tatctctctt	ttacgtaaag	4320
cctgcgagat	cctcttaagt	atagcgggga	agccatcggt	attcgatatt	gtcgttaacaa	4380
atactttgat	cggcgctatg	tttaaatgtt	taaacatgga	cagatatgcg	atgaaaacgc	4440
taagtgatac	tccaaatggt	gaaaggtagc	atgcttgga	acaatacttg	gaaatcaccg	4500
gaaacaccat	atgcggcgaa	aagccaatta	gtgtgatact	aagtgcctta	tcgaaaatcc	4560

-continued

gtgatgccgg	tccttcaggc	atcaaatctc	agtggcctaa	ttattcacag	agttctcatg	4620
tgacaagtat	tgatgatagt	agtgtcagtt	atgcttcagg	ttatgttact	ataggataat	4680
gatcacggct	aaaacggtcg	aatgtaagca	tatatctttc	gattgtataa	ttgttcccaa	4740
atactacagc	atctcaagga	aaaaaaaaa	aaaacttcca	aaaaaatcga	atccctgagg	4800
aatctttaat	acattttcaa	tctattttaag	ttttataaac	gtgtatatga	gatgtcatga	4860
gcatgaatta	ttaataataa	aaactaaatc	attaaagtaa	cttaaggagt	taaagcccg	4920
gctttaattg	ttagcagcct	tgacttgagc	aatcaattct	ggaacaacct	tgttgacatc	4980
accgacaatg	gccaaatcag	caaccttcat	gattggagct	tcgacatctt	tgttgatggc	5040
aatgatgtag	tcagagtctt	gcataccagc	caagtgttgg	atggcaccag	agataccaca	5100
agcaatgtac	aaagtgtgtc	tgacgggtct	accggtttga	cgcacttgca	agtccttgtc	5160
aaccatttcc	ttttcaatgg	cagctctgga	agcagcaatg	gtaccacca	acaagaagc	5220
taattcttcc	aatttttcga	agttttcctt	ggaaccaaca	ccacgaccac	cagcaaccaa	5280
aaccttggct	tcaccgatat	cagcaatgtc	cttggccaat	ttgacaacct	tggaaacctt	5340
ggttctgata	tcagaagcag	tcaatttgat	ggcaaccttt	tcgatcttgt	catcagaaac	5400
gttagcatcg	ttaactggca	attttttcaa	gacacctggg	ctgacgggtg	ccatttgagg	5460
tctgtggtca	gaacagacaa	tggtagcaat	caagttacca	ccgaaagctg	gtctggtagc	5520
caacaagtca	cggtttttga	catcgataat	caaagaggta	cagtcagcag	tcaaaccagt	5580
agacaatctg	gcagcaatc	ttggacccaa	gtctctaccg	atgaaagtag	caccgatgaa	5640
taagatttct	ggctttcttt	cgttgaccaa	gtcacagata	accttggcgt	aaccgtcagt	5700
ggagaaatga	gctaataatt	cgttgtcagc	agccaaaacc	ttgtcagcac	cgtgggacaa	5760
caagtcttgg	gacatctttt	cagtgttgtg	accaataaag	acagcagtc	attcaacacc	5820
caatttttca	gccatttctt	taccttacc	tagcaattcc	aaagaaacct	tttgaatttc	5880
accatctctt	tgttcagcga	aaaccagac	accctttag	tcagccttgt	tcattgttag	5940
ttaattatag	ttcgttgacc	gtatattcta	aaaacaagta	ctccttaaaa	aaaaaccttg	6000
aagggaataa	acaagtagaa	tagatagaga	gaaaaataga	aatgcaaga	gaatttatat	6060
attagaaaga	gagaaagaaa	aatggaaaaa	aaaaaatagg	aaaagccaga	aatagcacta	6120
gaaggagcga	caccagaaaa	gaaggtgatg	gaaccaatct	agctatatat	agttaactac	6180
cggctcgatc	atctctgcct	ccagcatagt	cgaagaagaa	tttttttttt	cttgaggctt	6240
ctgtcagcaa	ctcgtatttt	ttctttcttt	tttggtagag	ctaaaaagtt	cccacgttct	6300
cttgtacgac	gcggtcaca	acaaccttat	gggtaatttg	tcgcggtctg	ggtgtataaa	6360
tgtgtgggtg	caggccggcc	gtttaaacgg	gcgcgccacg	cgggtggagct	ccagcttttg	6420
ttccctttag	tgagggttaa	ttgcgcgctt	ggcgtaatca	tggtcatagc	tgtttcctgt	6480
gtgaaattgt	tatccgctca	caattocaca	caacatagga	gccggaagca	taaagtgtaa	6540
agcctggggg	gcctaagtga	tgaggtaact	cacattaatt	gcgttgcgct	cactgcccgc	6600
tttcagtcg	ggaaacctgt	cgtgccagct	gcattaatga	atcgccaac	gcgcggggag	6660
aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	6720
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	6780
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	6840

-continued

taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	6900
aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	6960
tccccctgga	agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	cgggatacct	7020
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	7080
cagttcgggtg	taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	7140
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	aaccggtaa	gacacgactt	7200
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	7260
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	7320
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	7380
acaaaccacc	gctggtagcg	gtggtttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	7440
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	7500
aaactcacgt	taagggatct	tgggtcatgag	attatcaaaa	aggatcttca	cctagatcct	7560
tttaaatata	aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttggtctga	7620
cagttacca	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcac	7680
catagttgcc	tgactccccg	tgcgttagat	aactacgata	cgggagggct	taccatctgg	7740
ccccagtgtc	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	7800
aaaccagcca	gccggaagg	ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	7860
ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	7920
caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	7980
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	8040
agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	8100
actcatgggt	atggcagcac	tgcataatc	tcttactgtc	atgccatccg	taagatgctt	8160
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	8220
ttgtctttgc	ccggcgctca	tacgggataa	taccgcgcca	catagcagaa	ctttaaaggt	8280
gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgtgttgag	8340
atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	8400
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	8460
gacacggaaa	tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	8520
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	8580
ggttccgcgc	acatttcccc	gaaaagtgcc	acctgacgtc	taagaaacca	ttattatcat	8640
gacattaacc	tataaaaata	ggcgtatcac	gaggcccttt	cgtc		8684

<210> SEQ ID NO 44

<211> LENGTH: 12314

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: pBOL115

<400> SEQUENCE: 44

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120

-continued

ttggcgggtg	tcggggcttg	cttaactatg	cggcacacaga	gcagattgta	ctgagagtgc	180
accataccac	agcttttcaa	ttcaattcat	catttttttt	ttattctttt	ttttgatttc	240
ggtttctttg	aaattttttt	gattcggtaa	tctccgaaca	gaaggaagaa	cgaaggaagg	300
agcacagact	tagattggta	tatatacgca	tatgtagtgt	tgaagaaaca	tgaattgcc	360
cagtattctt	aacccaactg	cacagaacaa	aaacctgcag	gaaacgaaga	taaatcatgt	420
cgaagctac	atataaggaa	cgtgctgcta	ctcatcctag	tcctgttgct	gccaaagctat	480
ttaatatcat	gcacgaaaag	caaacaaact	tgtgtgcttc	attggatggt	cgtaccacca	540
aggaattact	ggagttagtt	gaagcattag	gtcccaaat	ttgtttacta	aaaacacatg	600
tggatatctt	gactgatttt	tccatggagg	gcacagttaa	gccgctaaag	gcattatccg	660
ccaagtacaa	ttttttactc	ttcgaagaca	gaaaatttgc	tgacattggg	aatacagtca	720
aattgcagta	ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	780
acggtgtggt	gggccagggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	840
aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggtcc	ctatctactg	900
gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	960
ttattgctca	aagagacatg	ggtggaagag	atgaaggtta	cgattgggtg	attatgacac	1020
ccggtgtggg	tttagatgac	aaggagacg	cattgggtca	acagtataga	accgtggatg	1080
atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactattt	gcaaagggaa	1140
gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	1200
gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	1260
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tcgggtgtga	aataccgcac	1320
agatgcgtaa	ggagaaaata	ccgcacagc	aaattgtaaa	cgtaaatatt	ttgttaaaat	1380
tcgctgtaaa	tttttgtaa	atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	1440
tcctttataa	atcaaaagaa	tagaccgaga	tagggttgag	tgttggtcca	gtttggaaca	1500
agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	1560
gcgatggccc	actacgtgaa	ccatcacct	aatcaagttt	tttgggtcg	agggtccgta	1620
aagcactaaa	tcggaaccct	aaaggagacc	cccgatttag	agcttgacgg	ggaaagccgg	1680
cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaggagc	gggcgctagg	gcgctggcaa	1740
gtgtagcggg	cacgtgcgc	gtaaccacca	caccgcgcgc	gcttaatgcg	ccgctacagg	1800
gcgcgtcgcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaaggcgga	tcggtgcggg	1860
cctcttcgct	attacgccag	ctggcgaaag	gggatgtgc	tgcaaggcga	ttaagttggg	1920
taacgccagg	gttttccag	tcacgacgtt	gtaaaacgac	ggccagttag	cgcgcgtaat	1980
acgactcact	ataggcgcaa	ttgggtaccg	ggccccct	cgaggctgcg	ggtatcgata	2040
agcttgatat	cgaattcctg	cagccggggg	gatccactag	ttctagagcg	gccatttaa	2100
acggccggcc	ctagatcaga	gggtggtaaa	tgaagtgtaa	tagtattcat	ttttcttata	2160
aatcatccct	tccgtgattt	atacaaaaga	agaggagaat	atgctgaata	cttggtatat	2220
tactctacat	tatactctta	tcttgacggg	tattctgagc	atcttactca	gtttcaagat	2280
cttttaatgt	ccaaaaacat	ttgagccgat	ctaaatactt	ctgtgttttc	attaatttat	2340
aaattgtact	cttttaagac	atggaaagta	ccaacatcgg	tgaaacagt	ttttcattta	2400

-continued

cttatgggttt attgggtttt ccagtgaatg attatattgtc gttacccttt cgtaaaagtt	2460
caaacacggtt ttttaagtatt gtttagttgc tctttcgaca tatatgatta tccctgcgcg	2520
gctaaagtta aggatgcaaa aaacataaga caactgaagt taatttacgt caattaagtt	2580
ttccagggta atgatgtttt gggcttcac taattcaata agtatgtcat gaaatacgtt	2640
gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tgttgtttct	2700
tttctttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc	2760
aacgcagcta cggctaacaa agcctagtgg aactcgactg atgtgttagg gcctaaaact	2820
ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaagacct	2880
tcgagaccgc acgtacgaac acatacgtat gctaatatgt gttttgatag taccagtg	2940
tcgcagacct gcaattttt tgtaggttg gaagaatata taaaggttgc actcattcaa	3000
gatagttttt tcttgtgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaa	3060
cacaagaactt agtttcaaat taaattcatc acacaaacaa acaaaacaaa atgaacattg	3120
ttgtttgttt gaagcaagtt ccagacactg ctgaagtcag aattgacca gtcaagggt	3180
ctttaatcag agaaggtgtt ccattctatc tcaaccaga cgacaagaac gctttggaag	3240
aagcttttgg tttgaaggac aactacggtg ctacggttac cgtcatttcc atgggtccac	3300
ctcaagccaa gaacgctttg gttgaagctt tggccatggg tgctgatgaa gctgtcttat	3360
tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg	3420
gtatcaagaa attgaaatc gatatcgtct ttgccggtcg tcaagccatc gatggtgata	3480
ccgctcaagt cgggtccaga attgtgaac atttgggtat tccacaagtc acctacgttg	3540
aaaaggttga agttgacggt gacactttga agatcagaaa ggcttgggaa gacggttacg	3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa	3660
gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg	3720
ctgatgatat cgatgtgcac aaggccaact tgggtttgaa aggttctcca accaagggtca	3780
agaaatcttc taccaggaa gtcaagggtc aaggtgaagt cattgacaaa ccagtcaagg	3840
aagctgccgc ttacgttgtt tccaagttga aggaagaaca ctacatctaa agcccggtcg	3900
gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta	3960
attttttga atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct	4020
aaaaattccc ctcttttttc agttatatct taacaggcga cagtcacaaat gttgatttat	4080
cccagtcga ttcacaggg ttgtgaagca ttttgtcaat ggtcgaaatc acatcagtaa	4140
tagtgccctc tacttgccct atagaatttc tttctcttaa cgtcacggtt tggctcttta	4200
tagtttcgaa atctatggtg ataccaatg gtgttcccaa ttcacgtta cgggcgtatt	4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaaag	4320
cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa	4380
atactttgat cggcgctatg tttaaatgtt taaacatgga cagatatgcg atgaaaacgc	4440
taagtgatac tccaaatggt gaaaggtacg atgcttgga acaatacttg gaaatcacg	4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcctta tcgaaaatcc	4560
gtgatgccg tccttcagc atcaaattc agtggcctaa ttattcacag agttctcatg	4620
tgacaagtat tgatgatagt agtgctcagtt atgcttcagg ttatgttact ataggataat	4680

-continued

gatacaggct	aaaacggctg	aatgtaagca	tatatctttc	gattgtataa	ttgttcccaa	4740
atactacagc	atctcaagga	aaaaaaaaca	aaaacttcca	aaaaaatcga	atccctgagg	4800
aatctttaat	acattttcaa	tctattttaag	ttttataaac	gtgtatatga	gatgtcatga	4860
gcatgaatta	ttaataataa	aaactaaatc	attaaagtaa	cttaaggagt	taaagcccg	4920
gctttaattg	ttagcagcct	tgacttgagc	aatcaattct	ggaacaacct	tgttgacatc	4980
accgacaatg	gccaaatcag	caaccttcat	gattggagct	tcgacatctt	tgttgatggc	5040
aatgatgtag	tcagagtctt	gcataccagc	caagtgttgg	atggcaccag	agataccaca	5100
agcaatgtac	aaagttggtc	tgacggctct	accggtttga	ccgacttgca	agtccttgtc	5160
aaccatttcc	ttttcaatgg	cagctctgga	agcagcaatg	gtaccacca	acaaagaagc	5220
taattcttcc	aatttttcga	agttttcctt	ggaaccaaca	ccacgaccac	cagcaaccaa	5280
aaccttggct	tcaccgatat	cagcaatgtc	cttggccaat	ttgacaacct	tggaacctt	5340
ggttctgata	tcagaagcag	tcaatttgat	ggcaaccttt	tcgatcttgt	catcagaaac	5400
gttagcatcg	ttaactggca	atttttcaaa	gacacctggg	ctgacgggtg	ccatttgagg	5460
tctgtggtca	gaacagacaa	tggtagcaat	caagttacca	ccgaaagctg	gtctggtagc	5520
caacaagtca	cggttttcga	catcgatatc	caaagaggta	cagtcagcag	tcaaaccagt	5580
agacaatctg	gcagcaatc	ttggaccaa	gtctctaccg	atgaaagtag	caccgatgaa	5640
taagatttct	ggctttcttt	cgttgaccaa	gtcacagata	accttggcgt	aaccgtcagt	5700
ggagaaatga	gctaataatt	cgttgtcagc	agccaaaacc	ttgtcagcac	cgtgggacaa	5760
caagtccttg	gacatctttt	cagtgttggtg	acccaataag	acagcagtc	attcaacacc	5820
caatttttca	gccatttctt	taccttacc	tagcaattcc	aaagaaacct	tttgaattc	5880
accatctctt	tgttcagcga	aaaccagac	accctttag	tcagccttgt	tcattgttag	5940
ttaattatag	ttcgttgacc	gtatattcta	aaaacaagta	ctccttaaaa	aaaaaccttg	6000
aagggaataa	acaagtagaa	tagatagaga	gaaaaataga	aatgcaaga	gaatttatat	6060
attagaaaga	gagaaagaaa	aatggaaaaa	aaaaaatagg	aaaagccaga	aatagcacta	6120
gaaggagcga	caccagaaaa	gaaggtgatg	gaaccaattt	agctatatat	agttaactac	6180
cggctcgatc	atctctgcct	ccagcatagt	cgaagaagaa	tttttttttt	cttgaggctt	6240
ctgtcagcaa	ctcgtatttt	ttctttcttt	tttggtgagc	ctaaaaagtt	cccacgttct	6300
ctgtacgac	gccgtcacia	acaaccttat	gggtaatttg	tcgcggtctg	ggtgtataaa	6360
tgtgtgggtg	caggccggcc	gtttaaacgg	gccgccaccg	cgggtggagc	tgtgtggaag	6420
aacgattaca	acaggtgttg	tcctctgagg	acataaaaata	cacaccgaga	ttcatcaact	6480
cattgctgga	gttagcatat	ctacaattgg	gtgaaatggg	gagcgatttg	caggcatttg	6540
ctcggcatgc	cggtagaggt	gtggtcaata	agagcgacct	catgctatac	ctgagaaagc	6600
aacctgacct	acaggaaaga	gttactcaag	aataagaatt	ttcgttttaa	aacctaaag	6660
tcactttaaa	atttgtatata	acttattttt	tttataactt	atttaataat	aaaaatcata	6720
aatcataaga	aattcgctcg	agtcgactgc	agtttactgc	ttgtagtcgt	aagaagtttg	6780
gatgatatac	ttgatttcag	aaatcaaaagc	ttccttttggg	ttggcagtg	tacattggtc	6840
ttcgaaagcc	aattcagcca	ttctgtcaat	ggattcgttc	aattcttctt	cagagacacc	6900
ttgagatttc	aagttcattt	caataccaac	agattgacct	aattcgtaga	cagccttggc	6960

-continued

caaagattca	accaaagctt	cagtgggtgtt	acctttcaaa	cctaagaact	tggcgatata	7020
agcgtaatcg	gtgtcagctc	tgaagaattc	gtactttggg	aacaaagcgt	gcttttgagg	7080
gtccttggcg	ttgtatctga	tgatgtgagg	caacaagatg	gcgttagctc	taccatgtgg	7140
aataccgtat	tcaccaccaa	ttttgtgagc	aatggagtga	gcaataccca	agaaagcgtt	7200
agcaaaggcc	ataccggcca	aagtagaagc	gttgtgcata	ttttctctgg	aaacctgttc	7260
acctttttca	acggaagatt	tcaagtattc	aaaggtcaat	ttgatagctt	gtagggacaa	7320
acctctgggt	taatcggagg	ccatgacaga	aacgtaagat	tccatagcgt	gagtcaaaac	7380
gtccataccg	gtatcagcag	tgacagattt	tgggacggac	atgacaaatt	gaggggtcaat	7440
gatggcgaca	tctggagtca	aagcgaaatc	agccaatggg	tatttgacgt	tggtttcaga	7500
gtcagtata	acagcaaagt	gagtgacttc	agaaccagta	ccagaagtgg	ttggaataca	7560
gatgaaagt	gcgttttctg	gcataccaat	cttgtagggt	ctcttacc	tgccaagaa	7620
tttttgcttg	gcaccgaaga	aagaagtctc	tggatgttcg	aagaacatcc	acatagcctt	7680
ggcagcatcc	atggcagaac	caccacccaa	agcaatgatg	gtgtctgggt	ggaatcgac	7740
catcatttcc	aaaccttgt	agacagtgtt	ggtggatgga	tttggttcaa	cttcagagaa	7800
gatcttgatt	tgaggtgtt	cagttctttg	acgtaagacg	ttttcgacgg	tttggtgta	7860
accaaattca	accatacctg	ggtcacaaac	gatcatgacc	ttttcaatct	tgccatggt	7920
ggccaaggac	atgatagcgt	tttcttcgaa	atagatttga	gctggaacct	tgaagatttg	7980
agtgttggtt	cttctcttgg	caatgggtctt	gatgttcaac	aatcggttag	cagagacgtt	8040
gtgggagatg	gagtttctac	cgtaagaacc	acaacccaaa	gtcaaggatg	gaatcaattc	8100
gttgtacatg	tcaccgatac	caccaacagc	agatggagtg	ttgacaaaa	cacgacaagc	8160
cttcatttct	agaccgaaat	ccttttgcaa	agtttctgtc	tcagtgtgga	taacagcagt	8220
gtgacctaaa	ccaccgaagt	gcaaagtgtc	ttcacagatt	tggaaagctt	gcttggtaga	8280
ttgagccttg	actaaagcca	aaactggaga	caacttttct	ctggacaatg	ggtagtcaga	8340
accgacaccg	gagatttcag	caatgatcaa	tttggtgttt	tctggaactg	ggataccggc	8400
caattcagca	atttcaacag	cagacttacc	aacgatatct	ggcttgatac	cagtcttttg	8460
ttcgttcatg	atggcgtttt	ctaactcttg	taattcgtcc	ttcttgacga	agtaagcttg	8520
gtgagccttg	aattcattgg	tgacatcctt	gtagatttcc	ttgtcaatga	caacaacttg	8580
ttcagaagca	cagatcatac	cattatcgaa	agtcttgtaa	ccgatgatat	cgttgacagc	8640
acgcttgata	tgagcagctc	tttcgatgta	agatggaacg	ttacctggac	caacacccaa	8700
agctggttta	ccagtggagt	aagcagactt	aaccatacca	gaaccacggg	tagccaagac	8760
taaagcaata	cccttggtgt	tcataaattg	cttggttagct	tcaatggatg	gaacttcaat	8820
ccattggatg	atatcctttg	gagcaccagc	cttcatggca	gcttcacaaa	caacttcagc	8880
agctctcttg	gaagattctt	gagcagatgg	gtgaaaagcg	aaaataattg	ggttaccagt	8940
cttgatggca	atcatagcct	tgaagatggt	ggtagaagtt	gggttggtgg	ttggagtgc	9000
accacagatg	acaccaattg	gttcagcaac	gtaggtcaaa	cccttttctt	tgtcttcacc	9060
aatgatacca	acagtcttgg	tgctcttgat	ggagttccag	atgtattcag	aggcgtataa	9120
gttcttgata	gccttgctct	cgtagatacc	tctaccggtt	tcttcatgag	ccaacttggc	9180
caaaaccatg	tgttggtcaa	cagcagccaa	ggacatttgg	tggacaatgt	ggtcaatttc	9240

-continued

ttcttgagac	ttcttgagac	aagcttccaa	agccttctta	cctttgtcag	ctaaagcatc	9300
aatcatgatg	gcaacttctt	gttccttgga	acctctgttt	tccttttctg	gaatgggtcaa	9360
cattttttac	tagttctaga	atccgtcgaa	actaagttct	gggtgtttta	aactaaaaaa	9420
aagactaact	ataaaagtag	aatttaagaa	gtttaagaaa	tagatttaca	gaattacaat	9480
caatacctac	cgtctttata	tacttattag	tcaagtaggg	gaataatttc	agggaactgg	9540
tttcaacctt	ttttttcagc	tttttccaaa	tcagagagag	cagaaggtaa	tagaagggtg	9600
aagaaaaatga	gatagataca	tgcgtgggtc	aattgccttg	tgtcatcatt	tactccaggc	9660
agggtgcatc	actccattga	ggttgtgccc	gttttttgcc	tgtttgtgcc	cctgttctct	9720
gtagtgcgc	taagagaatg	gacctatgaa	ctgatggttg	gtgaagaaaa	caatattttg	9780
gtgctgggat	tctttttttt	tctggatgcc	agcttaaaaa	gcgggctcca	ttatatttag	9840
tggatgccag	gaataaactg	ttcaccacaga	cacctacgat	gttatatatt	ctgtgtaacc	9900
cgcctccctat	tttgggcatg	tacgggttac	agcagaatta	aaaggcta	tttttgacta	9960
aataaagtta	ggaaaatcac	tactattaat	tatttacgta	ttctttgaaa	tggcgagtat	10020
tgataatgat	aaactgagct	ccagcttttg	ttccctttag	tgagggttaa	ttgcgcgctt	10080
ggcgtaatca	tgggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	10140
caacatagga	gccggaagca	taaagtgtaa	agcctggggg	gcctaagtag	tgaggtaact	10200
cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagct	10260
gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	cgtattgggc	gctcttccgc	10320
ttctctcgtc	actgactcgc	tgcgtcgggt	cgttcggctg	cgccgagcgg	tatcagctca	10380
ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	10440
agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	10500
taggtctccg	ccccctgacg	agcatcacaa	aatcgacgc	tcaagtcaga	gggtggcgaaa	10560
cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	10620
tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	10680
gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	taggtcgttc	gctccaagct	10740
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	10800
tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	10860
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggg	ggcctaacta	10920
cggctacact	agaagacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	10980
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	11040
tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	11100
ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	taagggattt	tggtcattgag	11160
attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaaataaat	11220
ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	tgcttaataca	gtgaggcacc	11280
tatctcagcg	atctgtctat	ttcgttcctc	catagttgcc	tgactccccg	tcgtgtagat	11340
aactacgata	cgggaggggc	taccatctgg	ccccagtgct	gcaatgatac	cgcgagaccc	11400
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgag	11460
aagtgtcctc	gcaactttat	ccgcctccat	ccagtcctatt	aattgttgcc	gggaagctag	11520

-continued

```

agtaagtagt tgcgcagtta atagtttgcg caacggttgtt gccattgcta caggcatcgt 11580
gggtgcacgc tcgtcgtttg gtatggcttc attcagctcc gggtcccaac gatcaaggcg 11640
agttacatga tccccatgt tgtgcaaaaa agcgggttagc tccttcgggc ctccgacgt 11700
tgtcagaagt aagttggcgc cagtgttatc actcatggtt atggcagcac tgcataattc 11760
tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc 11820
attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc cggcgctcaa tacgggataa 11880
taccgcgcc catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt ctccggggcg 11940
aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca ctctgcacc 12000
caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag 12060
gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgtgaatac tcatactctt 12120
cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 12180
tgaatgtatt tagaaaaata aacaaatagg gggtccgcgc acatttcccc gaaaagtgcc 12240
acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgatcac 12300
gaggcccttt cgtc 12314

```

```

<210> SEQ ID NO 45
<211> LENGTH: 11180
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: pBOL116

```

```

<400> SEQUENCE: 45

```

```

tcgcgcgttt cgggtgatgac ggtgaaaacc tctgacacat gcagctcccc gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accataccac agcttttcaa ttcaattcat catttttttt ttattctttt ttttgatttc 240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg 300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaattgcc 360
cagtattctt aaccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt 420
cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct gccaaagctat 480
ttaatatcat gcacgaaaag caaacaact tgtgtgcttc attggtggt cgtaccacca 540
aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta aaaacacatg 600
tggaatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg 660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca 720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac 780
acggtgtggt gggcccagggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa 840
aggaacctag aggccttttg atgttagcag aattgtcatg caagggtcc ctatctactg 900
gagaatacat taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct 960
ttattgtcga aagagacatg ggtggaagag atgaagggtta cgattgggtg attatgacac 1020
ccggtgtggg tttagatgac aaggagacg cattgggtca acagtataga accgtggatg 1080
atgtggtctc tacaggatct gacattatta ttgttgaag aggactattt gcaaagggaa 1140

```

-continued

gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	1200
gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	1260
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	1320
agatgcgtaa	ggagaaaata	ccgcatcagg	aaattgtaaa	cgtaaatatt	ttgttaaaat	1380
tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	1440
tccttataaa	atcaaaagaa	tagaccgaga	tagggttgag	tgttggtcca	gtttggaaca	1500
agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	1560
gcgatggccc	actacgtgaa	ccatcacccct	aatcaagttt	tttggggtcg	aggtgccgta	1620
aagcactaaa	tcggaaccct	aaaggagacc	cccgathtag	agcttgacgg	ggaagccgg	1680
cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaggagc	gggcgctagg	gcgctggcaa	1740
gtgtagcggg	cacgctgcgc	gtaaccacca	caccgcgcgc	gcttaatgcg	ccgctacagg	1800
gcgcgtcgcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaagggcga	tcgggtcggg	1860
cctcttcgct	attacgccag	ctggcgaaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	1920
taacgccagg	gttttccag	tcacgacgtt	gtaaaacgac	ggccagtgag	cgcgcgtaat	1980
acgactcact	ataggcgcaa	ttgggtaccg	ggccccccct	cgaggctgac	ggtatcgata	2040
agcttgatat	cgaattcctg	cagccccggg	gatccactag	ttctagagcg	gcccatttaa	2100
acggccggcc	ctagatcaga	gggtggtaaa	tgaagtgtaa	tagtattcat	ttttcttata	2160
aatcatccct	tccgtgattt	atacaaaaga	agaggagaat	atgctgaata	cttggtatat	2220
tactctacat	tatactctta	tcttgacggg	tattctgagc	atcttactca	gtttcaagat	2280
cttttaaatgt	ccaaaaacat	ttgagccgat	ctaaatactt	ctgtgttttc	attaatttat	2340
aaattgtact	cttttaagac	atggaaaagta	ccaacatcgg	ttgaaacagt	ttttcattta	2400
cttatggttt	attggttttt	ccagtgaatg	attatttgtc	gttacccttt	cgtaaaagtt	2460
caaacacggt	tttaagtatt	gtttagtgtc	tctttcgaca	tatatgatta	tcctcgcgcg	2520
gctaaagtta	aggatgcaaa	aaacataaga	caactgaagt	taatttacgt	caattaagtt	2580
ttccagggtta	atgatgtttt	gggcttcac	taattcaata	agtatgtcat	gaaatacgtt	2640
gtgaagagca	tccagaaata	atgaaaagaa	acaacgaac	tgggtcggcc	tgttgtttct	2700
tttctttacc	acgtgatctg	cggcatttac	aggaagtcgc	gcgttttgcg	cagttgttgc	2760
aacgcagcta	cggctaacaa	agcctagtgg	aactcgactg	atgtgttagg	gcctaaaact	2820
gggtgtgaca	gctgaagtga	actattcaat	ccaatcatgt	catggctgtc	acaaagacct	2880
tgcgaccgc	acgtacgaac	acatacgtat	gctaatatgt	gttttgatag	taccagtgta	2940
tcgcagacct	gcaatttttt	tgtaggtttg	gaagaatata	taaagggtgc	actcattcaa	3000
gatagttttt	ttcttggtgt	tctattcatt	ttattattgt	ttgtttaaat	gttaaaaaaa	3060
ccaagaactt	agtttcaaat	taaattcatc	acacaaacaa	acaaaacaaa	atgaacattg	3120
ttgtttgttt	gaagcaagtt	ccagacactg	ctgaagtcag	aattgacca	gtcaagggtta	3180
ctttaatcag	agaagggtgt	ccatctatca	tcaaccaga	cgacaagaac	gctttggaag	3240
aagcttttgt	tttgaaggac	aactacgggtg	ctcacgttac	cgtcatttcc	atgggtccac	3300
ctcaagccaa	gaacgctttg	gttgaagctt	tggccatggg	tgctgatgaa	gctgtcttat	3360
tgactgacag	agctttcggt	ggtgtgata	ctttagctac	ctctcacacc	attgctgctg	3420

-continued

gtatcaagaa attgaaatac gatatcgtct ttgccggtcg tcaagccatc gatggtgata	3480
ccgctcaagt cgggccagaa attgctgaac atttgggtat tccacaagtc acctacgttg	3540
aaaaggttga agttgacggt gacactttga agatcagaaa ggcttgggaa gacggttacg	3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa	3660
gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg	3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggcca	3780
agaaatcttc taccaggaa gtcaagggtc aaggtgaagt cattgacaaa ccagtcaagg	3840
aagctgccgc ttacgttgtt tccaagttga aggaagaaca ctacatctaa agcccgggcg	3900
gagattgata agacttttct agttgcatac cttttatatt taaatcttat ctattagtta	3960
atTTTTtga atttatctt atatatagtc tggttattct aaaatatcat ttcagtatct	4020
aaaaattccc ctcttttttc agttatatct taacaggcga cagtcctaat gttgatttat	4080
cccagtccga ttcacagggt ttgtgaagca tttgtcaat ggtcgaaatc acatcagtaa	4140
tagtgccctc tacttgccct atagaatttc tttctcttaa cgtcacggtt tggctcttta	4200
tagtttcgaa atctatgggtg ataccaaagtg gtgttcccaa ttcacgttta cgggcgtatt	4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag	4320
cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa	4380
atactttgat cggcgctatg tttaaatgtt taaacatgga cagatatgcg atgaaaacgc	4440
taagtgatac tccaaatggt gaaaggtagc atgcttgaa acaatacttg gaaatcacgc	4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgtctta tcgaaaatcc	4560
gtgatgccgc tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg	4620
tgacaagtat tgatgatagt agtgcagtt atgcttcagg ttatgttact ataggataat	4680
gatcacggct aaaacggtcg aatgtaagca tatatcttcc gattgtataa ttgttcccaa	4740
atactacagc atctcaagga aaaaaaaca aaaacttcca aaaaaatcga atccctgagg	4800
aatctttaat acattttcaa tctattttaag ttttataaac gtgtatatga gatgtcatga	4860
gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaaggagt taaagcccg	4920
gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc	4980
accgacaatg gccaaatcag caaccttcac gattggagct tcgacatctt tgttgatggc	5040
aatgatgtag tcagagtctt gcataccagc caagtgttgg atggcaccag agataccaca	5100
agcaatgtac aaagttggtc tgacggctct accggtttga ccgacttgca agtccttgtc	5160
aaccatttcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaaagaagc	5220
taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaaccaa	5280
aaccttggct tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt	5340
ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac	5400
gttagcatcg ttaactggca atttttcaaa gacacctggt ctgacggtgg ccatttgagg	5460
tctgtggtca gaacagacaa tggtagcaat caagttacca ccgaaagctg gtctggtagc	5520
caacaagtca cggttttcga catcgatata caaagaggta cagtcagcag tcaaaccagt	5580
agacaatctg gcagcaatc ttggacccaa gtctctaccg atgaaagtag caccgatgaa	5640
taagatttct ggctttcttt cgttgaccaa gtcacagata accttggcgt aaccgtcagt	5700

-continued

ggagaaatga gctaataatt cggtgtcagc agccaaaacc ttgtcagcac cgtgggacaa	5760
caagtccttg gacatctttt cagtgttttg acccaataag acagcagtc attcaacacc	5820
caatttttca gccatttctt tacctttacc tagcaattcc aaagaaacct tttgtaattc	5880
accatctctt tgttcagcga aaaccacagc accctttag tagccttgt tcatgtttag	5940
ttaattatag ttctgtgacc gtatatctta aaaacaagta ctcttataaa aaaaaccttg	6000
aaagggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat	6060
attagaaaga gagaaagaaa aatggaaaaa aaaaaatagg aaaagccaga aatagcacta	6120
gaaggagcga caccagaaaa gaaggtgatg gaaccaattt agctatatat agttaactac	6180
cggctcgatc atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt	6240
ctgtcagcaa ctctgatttt ttctttcttt tttggtgagc ctaaaaagtt cccacgttct	6300
ctgttacgac gccgtcacia acaaccttat gggtaatttg tcgcggtctg ggtgtataaa	6360
tgtgtgggtg caggccggcc gtttaaaccg gccgccaccg cgggtggagc tgtgtggaag	6420
aacgattaca acaggtgttg tcctctgagg acataaaata cacaccgaga ttcatacaact	6480
cattgctgga gttagcatat ctacaattgg gtgaaatggg gagcgatttg caggcatttg	6540
ctcggcatgc cggtagaggt gtggtcaata agagcgacct catgctatac ctgagaaagc	6600
aacctgacct acaggaaaga gttactcaag aataagaatt ttcgttttaa aacctaaag	6660
tcactttaaa atttgtatac acttattttt ttataaactt atttaataat aaaaatcata	6720
aatcataaga aattcgctcg agtcgactgc agtttacaag ttcaatttgg ccatgatggc	6780
cttaacaatg gcttgaacat ctctgttctt ttctggttca gctggagcag aagaagaagc	6840
agcaccgaca ccgaaagctt ctctgatttc ttcgacggtg gtggtaccgt aagcaacctt	6900
tctgatgttg aacaagtttt ctggaccaac gttgtcagag gtggcagaac caccaacagc	6960
accacaacct aaagtcaaag atgggactaa gttggttagc ccaccgatac caccacaaga	7020
acctggagag ttaacaaaaa ttctggaac aggccttctt aaagcaaatt ctctaattgat	7080
ttcttcgttt tgagagtgga tgatcaaagt gtgaccagaa ccttggttgt gcaataaagc	7140
caaagacttt tcacaagctt catgccagtc ttcgacggtg tagaaagcca agactggagc	7200
caatttttcc ttagcgtatg gattttttag agaacatcg gtttgttcgg atagtaagat	7260
aacagcatca gatggaatgg aaataccagc caatttgcc aaagcttga catccttacc	7320
aacgatggct ggggttgtag tacctgtggc acgtaataga atcttaccaa ccttttcaga	7380
ttcttcagca ttcaagaagt aaccttttg tctcttgat tcttcgatga tttcagcctt	7440
cttgacggtt tcagcaatga tggattgttc agaagcacag atgacaccgt tgcgaaagt	7500
gtcagaaccg ataaccttcc taacagcagc tggaaatgca gcagttcttt cgatgaaaca	7560
tggaccgtta cctggaccga caccgatggc tggagtacca gaggagtaag cagctctaac	7620
cataccttca ccaccgtag ccaagatcaa agcgggtgtc ttgttcttca tcaattcagc	7680
agtaacctta acggtcaaaa tggacataca ttggatcaaa ccatctggag caccagcttc	7740
aacagcagcc ttttgcata tcttaacggt ttcagtgatg gaacggacag cagttgggtg	7800
tggagagaag acaatggcgt taccagcctt caaagcaatc aagacctga aaatggcagt	7860
ggaagtggg ttggttagat gaatacaacc agcaatgaca cctaattgga cagcaatgtc	7920
aatcaatttc ttttcttgt ctcccttcaa gataccaacg gtcttcaaat ccttgatgta	7980

-continued

gtttagaca acaatggagg agaatttggt cttgatgacc ttgtcttccc atttaccgta	8040
accggtgtct tcgtaagcca atttggccaa tttgacagct tcaacttcag tagccttggc	8100
gatcttttcg atgacctgtg taacagcttc ttgggaaaag ttcttgaatt cagcttgagc	8160
cttcttggcc ttggcaatca aagttctaac ttcttggatg gattgcaaat ccttgtccat	8220
gatttccatt ttttactagt tctagaatcc gtcgaaacta agttctgggtg ttttaaaact	8280
aaaaaaaaga ctaactataa aagtagaatt taagaagttt aagaaataga tttacagaat	8340
tacaatcaat acctaccgtc tttatatact tattagtcaa gtaggggaat aatttcaggg	8400
aactggttcc aacctttttt ttcagctttt tccaaatcag agagagcaga aggtaataga	8460
agggtgaaga aaatgagata gatacatgcg tgggtcaatt gccttgtgtc atcatttact	8520
ccaggcagggt tgcctcactc cattgagggt gtgcccgttt tttgctgtt tgtgcccctg	8580
ttctctgtag ttgcgctaag agaatggacc tatgaactga tggttggtga agaaaacaat	8640
atthtgggtgc tgggattctt tttttttctg gatgccagct taaaaagcgg gctccattat	8700
atthtagtga tgccaggaat aaactgttca cccagacacc tacgatgtta tatattctgt	8760
gtaaccgcc ccctattttg ggcattgtacg gggtacagca gaattaaaag gctaattttt	8820
tgactaaata aagttaggaa aatcactact attaattatt tacgtattct ttgaaatggc	8880
gagtattgat aatgataaac tgagctccag cttttgttcc ctttagtgag ggtaattgc	8940
gcgcttggcg taatcatggt catagctgtt tctgtgtga aattgttatc cgtcacaat	9000
tccacacaac ataggagcgg gaagcataaa gtgtaaagcc tgggggcct aatgagtga	9060
gtaactcaca ttaattgctg tgcgtcact gccgcttcc cagtcgggaa acctgtcgtg	9120
ccagctgcat taatgaatcg gccaacgcgc ggggagaggc gggttgcgta ttgggcgctc	9180
ttcgccttc tcgctcactg actcgtcgcg ctccgtcgtt cggctgcggc gagcggtatc	9240
agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa	9300
catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgcgtggcgtt	9360
ttcccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa gtcagagggtg	9420
gcgaaccgcc acaggactat aaagatacca ggcgtttccc cctggaagct cctcgtgcg	9480
ctctcctgtt ccgacctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag	9540
cgtggcgctt tctcatagct cacgctgtg gtatctcagt tcggtgtagg tcgttcgctc	9600
caagctgggc tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa	9660
ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccaactggcag cagccactgg	9720
taacaggatt agcagagcga ggtatgtagg cgggtctaca gagttcttga agtgggtggc	9780
taactacggc tacactagaa ggacagtatt tggatatctgc gctctgctga agccagttac	9840
cttcggaaaa agagttggta gctcttgatc cggcaaaaa accaccgctg gtagcgggtg	9900
ttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt	9960
gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag ggattttgt	10020
catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttta	10080
atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct taatcagtga	10140
ggcaccatc tcagcgatct gtctatttcg ttcacccata gttgcctgac tccccgtcgt	10200
gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcg	10260

-continued

```

agaccacgc tcaccggctc cagatttatac agcaataaac cagccagccg gaagggccga 10320
gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga 10380
agctagagta agtagttcgc cagttaataag tttgcgcaac gttgttgcca ttgctacagg 10440
catcgtgggtg tcacgctcgt cgtttggtat ggcttcattc agctccgggt cccaacgatac 10500
aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc 10560
gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca 10620
taattctctt actgtcatgc catccgtaag atgcttttct gtgactgggt agtactcaac 10680
caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccg cgtaatacg 10740
ggataatacc ggcgccata gcagaacttt aaaagtgtc atcattggaa aacgttcttc 10800
ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcg 10860
tgaccccaac tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac 10920
aggaaggcaa aatgccgcaa aaaaggggaat aagggcgaca cggaaatgt gaatactcat 10980
actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 11040
catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat tccccgaaa 11100
agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg 11160
tatcacgagg ccctttcgtc 11180

```

```

<210> SEQ ID NO 46
<211> LENGTH: 11108
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: pBOL118

```

```

<400> SEQUENCE: 46

```

```

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accataccac agcttttcaa ttcaattcat cttttttttt ttattctttt ttttgatttc 240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg 300
agcacagact tagattggtat tatatacgca tatgtagtgt tgaagaaaca tgaattgcc 360
cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt 420
cgaaagctac atataaggaa cgtgtctgcta ctcatcctag tcctgttgct gccaaagctat 480
ttaatatcat gcacgaaaag caaacaactt tgtgtgcttc attggatgtt cgtaccacca 540
aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta aaaacacatg 600
tggaatactt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg 660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca 720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac 780
acgggtgtgt gggcccagggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa 840
aggaacctag aggccttttg atgttagcag aattgtcatg caagggtcc ctatctactg 900
gagaataatac taagggtact gttgacattg cgaagagcga caaagatttt gttatcggtc 960
ttattgtctc aagagacatg ggtggaagag atgaagggtta cgattggttg attatgacac 1020

```

-continued

ccggtgtggg	tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	1080
atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactatct	gcaaagggaa	1140
gggatgctaa	gtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	1200
gatgcgccca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	1260
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	1320
agatgcgtaa	ggagaaaata	cgcacacagg	aaattgtaaa	cgtaaatatt	ttgttaaaat	1380
tcgcggtaaa	tttttgtaa	atcagctcat	tttttaacca	ataggccgaa	atcggcacaaa	1440
tcctttataa	atcaaaagaa	tagaccgaga	tagggttgag	tggtgttcca	gtttggaaca	1500
agagtccact	attaaagaac	gtggactcca	acgtcaaaag	gcgaaaaacc	gtctatcagg	1560
gcgatggccc	actacgtgaa	ccatcacccct	aatcaagttt	tttggggtcg	aggtgccgta	1620
aagcactaaa	tcggaaccct	aaaggagacc	cccgatttag	agcttgacgg	ggaaagccgg	1680
cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaggagc	ggcgctagg	gcgctggcaa	1740
gtgtagcgg	cacgtgcgc	gtaaccacca	caccgcgcgc	gcttaatgcg	ccgctacagg	1800
gcgcgtcgcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaaggcgca	tcggtgcggg	1860
cctcttcgct	attacgccag	ctggcgaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	1920
taacgccagg	gttttccag	tcacgacgtt	gtaaaacgac	ggccagttag	cgcgcgtaat	1980
acgactcact	ataggcgcaa	ttgggtaccg	ggccccccct	cgaggctgac	ggtatcgata	2040
agcttgatat	cgaattccct	cagcccgggg	gatccactag	ttctagagcg	gccattttaa	2100
acggccggcc	ctagatcaga	gggtggtaaa	tgaagtgtaa	tagtattcat	ttttcttata	2160
aatcatccct	tccgtgattt	atacaaaaga	agaggagaat	atgctgaata	cttggtatat	2220
tactctacat	tatactctta	tcttgacggg	tattctgagc	atcttactca	gtttcaagat	2280
cttttaagt	ccaaaaacat	ttgagccgat	ctaaatactt	ctgtgttttc	attaatttat	2340
aaattgtact	cttttaagac	atggaaagta	ccaacatcgg	tgaaacagt	ttttcattta	2400
cttatggttt	attggttttt	ccagtgaatg	attatttgtc	gttacccttt	cgtaaaagtt	2460
caaacacggt	tttaagtatt	gtttagtgtc	tctttcgaca	tatatgatta	tccctgcgcg	2520
gctaaagtta	aggatgcaaa	aaacataaga	caactgaagt	taatttacgt	caattaagtt	2580
ttccagggta	atgatgtttt	gggcttcac	taattcaata	agtatgtcat	gaaatacgtt	2640
gtgaagagca	tccagaaata	atgaaaagaa	acaacgaaac	tgggtcggcc	tggtgtttct	2700
tttctttacc	acgtgatctg	cggcatttac	aggaagtcgc	gcgttttgcg	cagttgtgtc	2760
aacgcagcta	cggctaacaa	agcctagtgg	aactcgactg	atgtgttagg	gcctaaaact	2820
gggtgtgaca	gctgaagtga	actattcaat	ccaatcatgt	catggctgtc	acaagacct	2880
tgcggaccgc	acgtacgaac	acatacgtat	gctaatatgt	gttttgatag	taccagtgta	2940
tcgcagacct	gcaatttttt	tgtaggtttg	gaagaatata	taaaggttgc	actcattcaa	3000
gatagttttt	ttctgtgtgt	tctattcatt	ttattattgt	ttgtttaaat	gttaaaaaaa	3060
ccaagaactt	agtttcaaat	taaattcatc	acacaaacaa	acaaaacaaa	atgaacattg	3120
ttgtttgttt	gaagcaagtt	ccagacactg	ctgaagtcag	aattgaccca	gtcaagggtta	3180
ctttaatcag	agaaggtgtt	ccatctatca	tcaaccaga	cgacaagaac	gctttggaag	3240
aagcttttgt	tttgaaggac	aactacggtg	ctcacgttac	cgtcatttcc	atgggtccac	3300

-continued

ctcaagccaa gaacgctttg gttgaagctt tggccatggg tgctgatgaa gctgtcttat	3360
tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg	3420
gtatcaagaa attgaaatag gatatcgctt ttgccggctg tcaagccatc gatggtgata	3480
ccgctcaagt cgggtccagaa attgctgaac atttgggtat tccacaagtc acctacgttg	3540
aaaaggttga agttgacggg gacactttga agatcagaaa ggcttgggaa gacgggttacg	3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa	3660
gatacatgtc cgttgaaaag atcttcgggtg ctttcgacaa ggaagtcaag atgtggactg	3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaagggtca	3780
agaaatcttc taccaggaa gtcaagggtc aaggtgaagt cattgacaaa ccagtcaagg	3840
aagctgccgc ttacgttgtt tccaagttga aggaagaaca ctacatctaa agcccgggcg	3900
gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta	3960
attttttgta atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct	4020
aaaaattccc ctcttttttc agttatatct taacaggcga cagtcctaat gttgatttat	4080
cccagtcga ttcatcaggg ttgtgaagca ttttgtcaat ggtcgaaatc acatcagtaa	4140
tagtgctctt tacttgcttc atagaatttc tttctcttaa cgtcacggtt tggctcttta	4200
tagtttcgaa atctatggg ataccaaatg gtgttcccaa ttcacgtta cgggcgtatt	4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag	4320
cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa	4380
atactttgat cggcgctatg tttaaatggt taaacatgga cagatatgcg atgaaaacgc	4440
taagtgtac tccaaatggt gaaaggtacg atgcttgga acaatacttg gaaatcacgc	4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcctta tcgaaaatcc	4560
gtgatgccgg tccttcagcg atcaaattc agtggcctaa ttattcacag agttctcatg	4620
tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat	4680
gatcacggct aaaacggctg aatgtaagca tatacttttc gattgtataa ttgttcccaa	4740
atactacagc atctcaagga aaaaaaaca aaacttcca aaaaaatcga atccctgagg	4800
aatctttaat acattttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga	4860
gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaaggagt taaagcccg	4920
gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc	4980
accgacaatg gccaaatcag caaccttcat gattggagct tcgacatctt tgttgatggc	5040
aatgatgtag tcagagtctt gcataaccagc caagtgttgg atggcaccag agataccaca	5100
agcaatgtac aaagtgggtc tgacgggtct accggtttga ccgacttgca agtccttgtc	5160
aaccatttc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaagaagc	5220
taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaaccaa	5280
aacctgggtc tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt	5340
ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac	5400
gttagcatcg ttaactggca atttttcaaa gacacctggg ctgacgggtg ccatttgagg	5460
tctgtggtca gaacagacaa tggtagcaat caagttacca ccgaaagctg gtctggtagc	5520
caacaagtca cggttttcga catcgatata caaagaggta cagtcagcag tcaaacagct	5580

-continued

agacaatctg gcagcaattc ttggacccaa gtctctaccg atgaaagtag caccgatgaa	5640
taagatttct ggctttcttt cgttgaccaa gtcacagata accttgccgt aaccgtcagt	5700
ggagaaatga gctaataatt cgttgtcagc agccaaaacc ttgtcagcac cgtgggacaa	5760
caagtccttg gacatctttt cagtgttggtg acccaataag acagcagtc aattcaacacc	5820
caatttttca gccatttcct tacccttacc tagcaattcc aaagaaacct tttgtaattc	5880
accatctctt tgttcagcga aaaccagac acccttgtag tcagccttgt tcatgtttag	5940
ttaattatag ttcgttgacc gtatatctta aaaacaagta ctcttataaa aaaaaccttg	6000
aagggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat	6060
attagaaaga gagaaagaaa aatggaaaaa aaaaaatagg aaaagccaga aatagcacta	6120
gaaggagcga caccagaaaa gaaggtgatg gaaccaattt agctatatat agttaactac	6180
cggctcgatc atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt	6240
ctgtcagcaa ctcgattttt ttctttcttt tttggtgagc ctaaaaagtt cccacgttct	6300
ctgtacgac gccgtcacia acaaccttat gggtaatgtg tcgcggtctg ggtgtataaa	6360
tgtgtgggtg caggccggcc gtttaaaccg gccgccaccg cggtgagacc tgtgtggaag	6420
aacgattaca acaggtgttg tcctctgagg acataaaata cacaccgaga ttcattcaact	6480
cattgctgga gtttagcatat ctacaattgg gtgaaatggg gagcgatttg caggcatttg	6540
ctcgcatgc cggtagaggt gtggtcaata agagcgacct catgctatac ctgagaaagc	6600
aacctgacct acaggaaaga gttactcaag aataagaatt ttcgttttaa aacctagag	6660
tcactttaaa atttgtatac acttattttt ttataactt atttaataat aaaaatcata	6720
aatcataaga aattcgctcg agtcgactgc agtttaaca attctgaaag catcgaccaa	6780
aacacaacga cgtaatctga cgaaagttct ggacagaagt acaccttcac cagttggggt	6840
ggtgatggtc atggtggtcc aaccttcacc acccaaacc aaaccagcga tacatggacc	6900
gttcttgaca aagatggaag tgtcaatggc gttagccatt tggttcatgt ttcgatgtt	6960
tctggagtgc atggcagcag tgtgtgaca accaccttc aatttgacag ccaaagcaat	7020
agcgtcagca acgttagcaa cacggacaac tggttaagact ggcatcatca attcagtac	7080
agcaaatggg tgttcagcgg tggtttcgac gaataataat ctggtttctt gtggaacctt	7140
caaacccgat gcagcagcaa tcttaccagc atctctacca acccagctc tggagacggg	7200
accttacct ctttcacga tgttcttcaa caaaactggg tgcaattgtt gagcttggtc	7260
agcagtcac ttgacggcat gttgacctc catcaatctc atcaattcgt cagcaacgga	7320
gtcaacaaca atcaaaacct tttcgtcagc acagatgatg ttgttgatga aagaagcacc	7380
cttgacaatg gattgagcag ctctggccaa atcagcgggt tcacgacaa caacaggagg	7440
gttaccagca ccagcagcaa tcaatctctt gttggtgtgc tttctggcag cttcaacaac	7500
agcttcacca ccagtgcga ctaatagacc gatacctggg aacttgaata atctttgagc	7560
agtttcgata tctgggttgg caacagtgc caacaagttt tctggaccac cagcagcaac	7620
aatggcttgg ttcaatagag tgatggctct ttgagaaacc ttcttgccag ctgggtgtgg	7680
agcgaagata acggagttag cagcagcaat caaagagatg gcgttggtga tgacagtgc	7740
agctgggttg gtagatgggg tgacggaagc aacaacacc catggagcat tttcaatcaa	7800
agtcaaacca ttatcacccg tcaagacttg tggagacaaa cattcgacac ctggagtacc	7860

-continued

tctagcttga gcaacgttct tagcgaattt gtcttcaact ctaccatac cggtttcgga	7920
gacagccaat tcagccaagt ctctggcatg cttttcacca gcttctctga tggcagcaat	7980
ggccaattgt ctcatggcaa cagattttcaa accttggtga gcaaccttg cagcagcaac	8040
agcgtcgtcc aaagaagcga aaacacccat ttcgtggaca gcagcagatg gagtgtcaga	8100
agattgcatt ttcaacaaga cagccttgac aacttggtcg atatcttggt ggttcatttt	8160
ttactagtcc tagaatccgt cgaaactaag ttctgggtgt ttaaaactaa aaaaaagact	8220
aactataaaa gtagaattta agaagtttaa gaaatagatt tacagaatta caatcaatac	8280
ctaccgtctt tatatactta ttagtcaagt aggggaataa tttcaggga ctggtttcaa	8340
cctttttttt cagctttttc caaatcagag agagcagaag gtaatagaag gtgtaagaaa	8400
atgagataga tacatgcgtg ggtcaattgc cttgtgtcat catttactcc aggcaggttg	8460
catcaectca ttgaggttgt gcccgttttt tgcctggttg tgccctggt ctctgtagt	8520
gcgctaagag aatggaccta tgaactgatg gttggtgaag aaaacaatat tttggtgctg	8580
ggattctttt tttttctgga tgccagctta aaaagcgggc tccattatat ttagtggatg	8640
ccaggaataa actgttcacc cagacaccta cgatgttata tattctgtgt aacccgcccc	8700
ctattttggg catgtacggg ttacagcaga attaaaaggc taattttttg actaaataaa	8760
gttaggaaaa tcactactat taattattta cgtattcttt gaaatggcga gtattgataa	8820
tgataaactg agctccagct tttgttcctt ttagtgaggg ttaattgcgc gcttgcgta	8880
atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat	8940
aggagccgga agcataaagt gtaaagcctg ggggtgcctaa tgagtgaggt aactcacatt	9000
aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgctcgtgc agctgcatta	9060
atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcttc	9120
gctcactgac tcgctgcgct cggctcgttcg gctgcggcga gcggtatcag ctcactcaaa	9180
ggcggaataa cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa	9240
aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct	9300
ccgccccct gacgagcacc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac	9360
aggactataa agataaccagg cgtttccccc tggaagctcc ctcgtgcgct ctctgttcc	9420
gaccctgcgc cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc	9480
tcatagtcga cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg	9540
tgtgcacgaa cccccgttc agcccgaccg ctgcgcctta tccggtaact atcgtcttga	9600
gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta acaggattag	9660
cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtagccta actacggcta	9720
cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag	9780
agttggtagc tcttgatccg gcaaaaaaac caccgctggt agcgggtggt tttttgtttg	9840
caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac	9900
ggggtctgac gctcagtgga acgaaaaact acgttaaggg attttgggtca tgagattatc	9960
aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat caatctaaag	10020
tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg cacctatctc	10080
agcgatctgt ctatttcgtt catccatagt tgccctgact cccgtcgtgt agataactac	10140

-continued

gatacgggag	ggcttaccat	ctggccccag	tgtgtcaatg	ataccgcgag	acccacgctc	10200
accgggtcca	gatttatcag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	10260
tcttgcaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	ctagagtaag	10320
tagttcgcca	gttaaatggt	tgcgcaacgt	tgttgccatt	gctacaggca	tcgtggtgtc	10380
acgctcgctg	tttggtatgg	cttcattcag	ctccgggttc	caacgatcaa	ggcgagttac	10440
atgatcccc	atgttgtgca	aaaaagcgg	tagctccttc	ggctcctcga	tcgttgtcag	10500
aagtaagttg	ggcgagtggt	tatcactcat	ggttatggca	gcactgcata	attctcttac	10560
tgctatgcca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	10620
agaatagtgt	atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	ataataccgc	10680
gccacatagc	agaactttaa	aagtgtctat	cattggaaaa	cggtcttcgg	ggcgaaaaact	10740
ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	ccactcgtg	cacccaactg	10800
atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	10860
tgcgcgaaaa	aagggaaata	gggcgacacg	gaaatgttga	atactcatac	tcttcctttt	10920
tcaatattat	tgaagcattt	atcaggggta	ttgtctcatg	agcggataca	tatttgaatg	10980
tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	11040
cgtctaagaa	accattatta	tcatgacatt	aacctataaa	aataggcgta	tcacgaggcc	11100
ctttcgtc						11108

<210> SEQ ID NO 47

<211> LENGTH: 11114

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: pBOL120

<400> SEQUENCE: 47

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accataccac	agcttttcaa	ttcaattcat	catttttttt	ttattctttt	ttttgatttc	240
ggtttctttg	aaattttttt	gattcggtaa	tctccgaaca	gaaggaagaa	cgaaggaagg	300
agcacagact	tagattggta	tatatacgca	tatgtagtgt	tgaagaaaca	tgaattgcc	360
cagtattctt	aacccaactg	cacagaacaa	aaacctgcag	gaaacgaaga	taaatcatgt	420
cgaagctac	atataaggaa	cgtgtgcta	ctcatcctag	tcctgttgct	gccaaagctat	480
ttaatatcat	gcacgaaaag	caaacaaact	tgtgtgcttc	attggatggt	cgtaccacca	540
aggaattact	ggagttagtt	gaagcattag	gtcccaaaat	ttgtttacta	aaaacacatg	600
tggaatctt	gactgatttt	tccatggagg	gcacagttaa	gccgctaaag	gcattatccg	660
caaagtacaa	ttttttactc	ttcgaagaca	gaaaatttgc	tgacattggt	aatacagtca	720
aattgcagta	ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	780
acggtgtggt	gggccagggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	840
aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggtccc	ctatctactg	900
gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	960

-continued

ttattgctca aagagacatg ggtggaagag atgaaggta cgattggttg attatgacac	1020
ccggtgtggg ttagatgac aaggagacg cattgggtca acagtataga accgtggatg	1080
atgtggtctc tacagatct gacattatta ttgttggaag aggactatct gcaaagggaa	1140
gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctgggaagca tatttgagaa	1200
gatgcgccca gcaaaactaa aaaactgtat tataagtaaa tgcattgata ctaaaactac	1260
aaattagagc ttcaatttaa ttatatcagt tattacccta tgcggtgtga aataccgcac	1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgtaaatatt ttgttaaaat	1380
tcgcgtaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa	1440
tccttataa atcaaaagaa tagaccgaga tagggttgag tgtgttcca gtttggaaca	1500
agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg	1560
gcgatggccc actacgtgaa ccatcacct aatcaagttt tttgggtcg aggtgccgta	1620
aagcactaaa tcggaacct aaaggagcc ccgatttag agcttgacgg ggaagccgg	1680
cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa	1740
gtgtagcggc cacgctgcgc gtaaccacca caccgcgcgc gcttaatgcg ccgctacagg	1800
gcgctgcgc ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg	1860
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagtggg	1920
taacgccagg gttttccag tcacgacgtt gtaaaacgac ggccagttag ccgctgaat	1980
acgactcact ataggcgcaa ttgggtaccg ggccccct cgaggctgac ggtatcgata	2040
agcttgatat cgaattcctg cagccgggg gatccactag ttctagagcg gccatttaa	2100
acggccggcc ctagatcaga ggggtgtaaa tgaagtgtaa tagtattcat tttcttata	2160
aatcatccct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat	2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat	2280
cttttaatgt ccaaaaacat ttgagccgat ctaatactt ctgtgttttc attaatattat	2340
aaattgtact cttttaagac atggaaagta ccaacatcgg ttgaaacagt ttttcattta	2400
cttatggtt attggtttt ccagtgaatg attattgtc gttaccctt cgtaaaagt	2460
caaacacgtt ttaagtatt gtttagttgc ttttcgaca tatatgatta tcctgcgcg	2520
gctaaagtta aggatgcaaa aacataaga caactgaagt taatttacgt caattaagtt	2580
ttccagggtat atgatgtttt gggcttcac taattcaata agtatgtcat gaaatacgtt	2640
gtgaagagca tcagaaata atgaaaagaa acaacgaac tgggtcgcc tgtgtttct	2700
tttctttacc acgtgatctg cggcatctac aggaagtcgc gcgttttgcg cagttgttc	2760
aacgcagcta cggctaaca agcctagtgg aactcgactg atgtgttagg gcctaaaact	2820
gggtgtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct	2880
tcgggaccgc acgtacgaac acatacgtat gctaataatgt gttttgatag taccagtg	2940
tcgcagacct gcaattttt ttaggtttg gaagaatata taaaggttc actcattcaa	3000
gatagtttt tcttctgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaa	3060
ccaagaactt agtttcaa taaattcatc acacaaacaa acaaaacaaa atgaacattg	3120
ttgtttgtt gaagcaagtt ccagacactg ctgaagtcag aattgacca gtaagggt	3180
ctttaatcag agaaggtgtt ccatctatca tcaaccaga cgacaagaac gctttggaag	3240

-continued

aagcttttgg	tttgaaggac	aactacgggtg	ctcacgttac	cgtcatttcc	atgggtccac	3300
ctcaagccaa	gaacgctttg	gttgaagctt	tggccatggg	tgtgatgaa	gctgtcttat	3360
tgactgacag	agctttcggg	ggtgctgata	cttttagctac	ctctcacacc	attgctgctg	3420
gtatcaagaa	attgaaatac	gatatcgtct	ttgccggctg	tcaagccatc	gatggtgata	3480
ccgctcaagt	cggctccagaa	attgctgaac	atttgggtat	tccacaagtc	acctacgttg	3540
aaaaggttga	agttgacggg	gacactttga	agatcagaaa	ggcttgggaa	gacggttacg	3600
aagttgttga	agtcaagact	ccagttctat	tgactgccat	caaggaattg	aacgttccaa	3660
gatacatgtc	cgttgaaaag	atcttcgggtg	ctttcgacaa	ggaagtcaag	atgtggactg	3720
ctgatgatat	cgatgtcgac	aaggccaact	tgggtttgaa	aggttctcca	accaagggtca	3780
agaaatcttc	taccaaggaa	gtcaagggtc	aaggtgaagt	cattgacaaa	ccagtcaagg	3840
aagctgcccg	ttacgttgtt	tccaagttga	aggaagaaca	ctacatctaa	agccccggcg	3900
gagattgata	agacttttct	agttgcatat	cttttatatt	taaatcttat	ctattagtta	3960
attttttgta	atttatcctt	atatatagtc	tggttattct	aaaatatcat	ttcagtatct	4020
aaaaattccc	ctcttttttc	agttatatct	taacaggcga	cagtcctaat	gttgatttat	4080
cccagtcgga	ttcatcaggg	ttgtgaagca	ttttgtcaat	ggtcgaaatc	acatcagtaa	4140
tagtgccctc	tacttgccct	atagaatttc	tttctcttaa	cgtcaccgtt	tggctcttta	4200
tagtttcgaa	atctatgggtg	ataccaaagt	gtgttcccaa	ttcatcgtaa	cgggcgtatt	4260
ttttaccaat	tgaagtattg	gaatcgtcaa	ttttaagta	tatctctctt	ttacgtaaag	4320
cctgcgagat	cctcttaagt	atagcgggga	agccatcggt	attcgatatt	gtcgtaaaca	4380
atactttgat	cggcgctatg	tttaaatgtt	taaacatgga	cagatatgcg	atgaaaacgc	4440
taagtgatac	tccaaatggt	gaaaggtagc	atgcttgtaa	acaatacttg	gaaatcacgc	4500
gaaacaccat	atgcggcgaa	aagccaatta	gtgtgatact	aagtgtctta	tcgaaaatcc	4560
gtgatgccgg	tccttcaggc	atcaaatttc	agtggcctaa	ttattcacag	agttctcatg	4620
tgacaagtat	tgatgatagt	agtgtcagtt	atgcttcagg	ttatgttact	ataggataat	4680
gatcacggct	aaaacggctg	aatgtaagca	tatatcttcc	gattgtataa	ttgttcccaa	4740
atactacagc	atctcaagga	aaaaaaaaa	aaaacttcca	aaaaaatcga	atccctgagg	4800
aatctttaat	acattttcaa	tctattttaag	ttttataaac	gtgtatatga	gatgtcatga	4860
gcatgaatta	ttaataataa	aaactaaatc	attaaagtaa	cttaaggagt	taaagcccg	4920
gctttaattg	ttagcagcct	tgacttgagc	aatcaattct	ggaacaacct	tgttgacatc	4980
accgacaatg	gccaaatcag	caaccttcat	gattggagct	tcgacatctt	tgttgatggc	5040
aatgatgtag	tcagagtctt	gcataaccagc	caagtgttgg	atggcaccag	agataccaca	5100
agcaatgtac	aaagttggtc	tgacggctct	accggtttga	ccgacttgca	agtccttgtc	5160
aaccatttcc	ttttcaatgg	cagctctgga	agcagcaatg	gtaccacca	acaaagaagc	5220
taattcttcc	aatttttcga	agttttcctt	ggaaccaaca	ccacgaccac	cagcaaccaa	5280
aaccttggct	tcaccgatat	cagcaatgtc	cttggccaat	ttgacaacct	tggaaacctt	5340
ggttctgata	tcagaagcag	tcaatttgat	ggcaaccttt	tcgatcttgt	catcagaaac	5400
gttagcatcg	ttaactggca	atttttcaaa	gacacctggg	ctgacggtgg	ccatttgagg	5460
tctgtgggtca	gaacagacaa	tggtagcaat	caagttacca	ccgaaagctg	gtctggtagc	5520

-continued

caacaagtca	cggttttcga	catcgatatac	caaagaggta	cagtcagcag	tcaaaccagt	5580
agacaatctg	gcagcaattc	ttggacccaa	gtctctaccg	atgaaagtag	caccgatgaa	5640
taagatttct	ggctttcttt	cgttgaccaa	gtcacagata	accttgccgt	aaccgtcagt	5700
ggagaaatga	gctaataatt	cgttgtcagc	agccaaaacc	ttgtcagcac	cgtgggacaa	5760
caagtccttg	gacatctttt	cagtgtttgt	acccaataag	acagcagtca	attcaacacc	5820
caatttttca	gccatttcct	tacccttacc	tagcaattcc	aaagaaacct	tttgaatttc	5880
accatctctt	tgttcagcga	aaaccagac	acccttgtag	tcagccttgt	tcattgtttag	5940
ttaattatag	ttcgttgacc	gtatattcta	aaaacaagta	ctccttaaaa	aaaaaccttg	6000
aagggaataa	acaagtagaa	tagatagaga	gaaaaataga	aatgcaaga	gaatttatat	6060
attagaaaga	gagaaagaaa	aatggaaaaa	aaaaaatagg	aaaagccaga	aatagcacta	6120
gaaggagcga	caccagaaaa	gaaggtgatg	gaaccaattt	agctatatat	agttaactac	6180
cggctcgatc	atctctgcct	ccagcatagt	cgaagaagaa	tttttttttt	cttgaggctt	6240
ctgtcagcaa	ctcgtatttt	ttctttcttt	tttggtgagc	ctaaaaagtt	cccacgttct	6300
cttgtacgac	gccgtcacia	acaaccttat	gggtaatttg	tcgcggtctg	ggtgtataaa	6360
tgtgtgggtg	caggccggcc	gtttaaacgg	gccgccaccg	cggtgaggcc	tgtgtggaag	6420
aacgattaca	acagggtgtg	tcctctgagg	acataaaaata	cacaccgaga	ttcatcaact	6480
cattgtctga	gttagcatat	ctacaattgg	gtgaaatggg	gagcgatttg	caggcatttg	6540
ctcggcatgc	cggtagaggt	gtggtcataa	agagcgacct	catgctatac	ctgagaaagc	6600
aacctgacct	acaggaaaga	gttactcaag	aataagaatt	ttcgttttaa	aacctaaagag	6660
tcactttaaa	atttgtatac	acttattttt	tttataactt	atttaataat	aaaaatcata	6720
aatcataaga	aattcgctcg	agtcgactgc	agtttatcta	atggagaaac	catcagtcia	6780
gacacatctt	cttctcttag	cgaagtgcgc	ggcagtggtg	gtaccttcac	cagttggagt	6840
agcaatgggtg	aaagtgggtg	aaccttcacc	tctgaaacct	aaaccagcga	aagatggacc	6900
gttcttgaca	aagatggagg	tttgcattgc	acgggcagcc	ttgttcaatc	tggagatggt	6960
ttgagagtgc	atggtagcag	tgtgatgtag	accttggttc	aattcaatgg	caacttciaa	7020
agcttcatcg	aagtctggaa	ctctgacaa	tggacaatt	ggcatcaaca	attcaacagt	7080
agcgaatggg	tgggactttt	cagtttcgac	aatgatcaat	cttgggggtg	aatcacaagc	7140
aataccagct	tctttcaaga	tttcagtggc	agacttacca	accaatttct	tgttggtgac	7200
acctttgtca	gtgacggcaa	ccttttccaa	tttttgata	tcagatgggt	tagtgacgtg	7260
caaagcaccg	ttcttttcca	tttgaacaa	caagaagtca	gcaatggagt	caacggcaac	7320
aacagacttt	tcagcgatgc	acaagatatt	atggtcaaag	gaagcaccgt	cgacaatgtc	7380
agcagcagcc	ttttcaatgt	tagcggtttc	gtcaacgatg	gatggagggt	taccagcacc	7440
agcaccgata	accttcttac	cagattgcat	agcttgcaag	acaacacctg	gaccaccagt	7500
gatgaccaac	aatggaacct	ttgggtggtt	catcatttct	tgagcagctt	ggatagatgg	7560
cttggcaacg	gtgacaatca	agttgtcaat	accacaagaa	tctctgacga	tagtgttcaa	7620
cttttcaatc	aaccataaag	agatgttctt	ggcacctggg	tgaggagagt	agaaaacggc	7680
gttaccagca	gccaacatac	cgatggaggt	acagatcaaa	gtttcagttg	ggttggtaga	7740
tggagcaaca	gcaccgatga	caccgtatgg	agataattcg	tataaagtca	taccgttgct	7800

-continued

accggtagca	acttcagtgt	acaagtcttc	aacacctgga	gtcttttcga	tagctaaagt	7860
gttcttcaag	attttatcgg	tgacattacc	cataccggtt	tcagcaacag	ctctggtagc	7920
aatggtttcg	atttctgggt	ataaagcttc	tctgatggcc	ttgacaacgt	ttcttctttc	7980
ttccaaagat	ttttccttgt	aacagttttg	agcaatgacg	gcagcttga	cagcttcac	8040
gacggtatcg	aaaaacccg	acttggcacc	ttgggtggtg	gtcttggttg	gaacttcctt	8100
ttgttcagcc	aatttttcca	acaaaacctt	cttgactaat	tgttccaatt	ccaaagattc	8160
cattttttac	tagttctaga	atccgtcgaa	actaagttct	gggtgtttta	aactaaaaaa	8220
aagactaact	ataaaagtag	aatttaagaa	gtttaagaaa	tagatttaca	gaattacaat	8280
caatacctac	cgtctttata	tacttattag	tcaagtaggg	gaataatttc	agggaaactgg	8340
tttcaacctt	tttttccagc	tttttccaaa	tcagagagag	cagaaggtaa	tagaagggtg	8400
aagaaaatga	gatagataca	tgcggtgggc	aattgccttg	tgtcatcatt	tactccaggc	8460
aggttgcatc	actccattga	ggttggtgcc	gttttttgcc	tggttggtgc	cctgttctct	8520
gtagttgcgc	taagagaatg	gacctatgaa	ctgatggttg	gtgaagaaaa	caatattttg	8580
gtgctgggat	tctttttttt	tctggatgcc	agcttaaaaa	gcgggctcca	ttatatattag	8640
tggtatgcca	gaataaaactg	ttcaccacga	cacctacgat	gttatatatt	ctgtgtaacc	8700
cgtcccttat	tttgggcatg	tacgggttac	agcagaatta	aaaggctaata	tttttgacta	8760
aataaagtta	ggaaaatcac	tactattaat	tatttacgta	ttctttgaaa	tgccgagtat	8820
tgataatgat	aaactgagct	ccagcttttg	ttccctttag	tgagggttaa	ttgcgcgctt	8880
ggcgtaata	tggtcatagc	tgtttctgt	gtgaaattgt	tatccgctca	caattccaca	8940
caacatagga	gccggaagca	taaagtgtaa	agcctggggg	gcctaagtag	tgaggtaact	9000
cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaaacctgt	cgtgccagct	9060
gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	cgtattgggc	gctcttccgc	9120
ttctctgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	tatcagctca	9180
ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	9240
agcaaaaagg	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	9300
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaga	ggtggcgaaa	9360
cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	9420
tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcggtgg	9480
gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	taggtcgctc	gctccaagct	9540
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	9600
tcttgagtcc	aaccgggtaa	gacacgactt	atgccactg	gcagcagcca	ctggtaacag	9660
gattagcaga	gcgaggtagt	taggcggtgc	tacagagtgc	ttgaagtggg	ggcctaacta	9720
cggctacact	agaagacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	9780
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	9840
tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	9900
ttctacgggg	tctgacgctc	agtgaacga	aaactcacgt	taagggtatt	tggtcatgag	9960
attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaaatcaat	10020
ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaata	gtgaggcacc	10080

-continued

tatctcagcg	atctgtctat	tctgttcac	catagttgcc	tgactccccg	tctgttagat	10140
aactacgata	cgggaggggt	taccatctgg	cccagtgct	gcaatgatac	cgcgagaccc	10200
acgctcacgg	gtccagatt	tatcagcaat	aaaccagcca	gccggaaggg	cgcgagcgag	10260
aagtggctct	gcaactttat	ccgcctccat	ccagtcctatt	aattgttgcc	gggaagctag	10320
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	10380
ggtgtcacgc	tctgtcttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	10440
agttacatga	tcccccatgt	tgtgcaaaaa	agcgggttagc	tccttcgggc	ctccgatcgt	10500
tgtcagaagt	aagttggccg	cagtggtatc	actcatgggt	atggcagcac	tgcataattc	10560
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtagt	caaccaagtc	10620
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgctca	tacgggataa	10680
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	cttcggggcg	10740
aaaaactctc	aggatcttac	cgctgttgag	atccagttcg	atgtaacca	ctcgtgcacc	10800
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	10860
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatactctt	10920
cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	10980
tgaatgtatt	tagaaaaata	aacaaatagg	ggttcgcgcg	acatttcccc	gaaaagtgcc	11040
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	11100
gaggcccttt	cgtc					11114

<210> SEQ ID NO 48

<211> LENGTH: 2613

<212> TYPE: DNA

<213> ORGANISM: Entamoeba histolytica

<400> SEQUENCE: 48

atgtcaacac	aacaaactat	gactgtagat	gaacatatta	atcaacttgt	tgctaaagca	60
caagttgcac	ttaaagaata	tcttaaacca	gaatatacac	aagaaaaaat	agattatatt	120
gtaaagaaag	catcagttgc	agcacttgat	caacattgtg	cacttgacgc	agctgcagtt	180
gaagaaacag	gaagagggtat	ttttgaagat	aaagctacta	aaaatatatt	tgcatgtgaa	240
catgtttacac	atgaaatgag	acatgctaaa	acagttggta	ttattaatgt	agatccactt	300
tatggaatta	cagaaattgc	agaaccagtt	ggagttgttt	gtggagttac	accagttact	360
aatccaacat	caacagctat	tttcaagtca	cttattttcaa	ttaaaacaag	aaatccaatt	420
gtattttcat	tccatccatc	agcacttaaa	tgttctatta	tggcagctaa	aattgttaga	480
gatgcagcta	ttgcagcagg	agcaccagaa	aattgtattc	aatggattga	atttgaggga	540
attgaagcat	caaataaatt	aatgaatcat	ccaggagttg	ctactattct	tgctacagga	600
ggaaatgcta	tggttaaagc	agcatattca	tcaggaaaac	cagcacttgg	agtaggagca	660
ggaaatgtac	caacatatat	tgaaaaaaca	tgtaatatta	aacaagcagc	aatgatgta	720
gttatgtcaa	aatcattttga	taatgggtatg	atttgtgcat	cagaacaagc	agcaattatt	780
gataaagaaa	tttatgatca	agtagttgaa	gaaatgaaaa	cacttggagc	atatttcatt	840
aatgaagaag	aaaaagctaa	attagaaaag	tttatgtttg	gagttaatgc	atattcagca	900
gatgttaata	atgcaagact	taatccaaaa	tgtccaggta	tgtcaccaca	atggtttgct	960

-continued

gaacaagttg gaattaaagt tccagaagat tgtaatatta tttgtgcagt ttgtaaagaa	1020
gttggaccaa atgaaccatt aacaagagaa aaattatcac cagttcttgc tattcttaaa	1080
gcagaaaata cacaagatgg tattgataaa gctgaagcta tggttgaatt taatggtaga	1140
ggacattcag cagctattca ttcgaatgat aaagcagtag ttgaaaagta tgcacttaca	1200
atgaaagcat gcagaatttt acataatata ccatcatcac aaggaggaat tggatcaatt	1260
tataactata tttggccatc atttacctt ggatgtggat catatggagg aaattcggta	1320
tcagctaattg ttacatatca taatttatta aatattaaaa gacttgcaga tagaagaaac	1380
aaccttcaat ggttcagagt tccaccaaag attttctttg aaccacattc tattagatat	1440
cttgcagaac ttaaggaact tagtaaaata ttcattgttt cagatagaat gatgtataaa	1500
ttaggatatg tagatagagt tatggatgta ttgaaaagaa gaagtaatga agtagaaatt	1560
gaaattttca ttgatgtaga accagatcca tctattcaaa ccgttcaaaa aggacttgct	1620
gttatgaata catttggacc agataatatt attgctattg gaggaggatc agctatggat	1680
gcagctaaga ttatgtggtt actttatgaa catccagaag ccgatttctt tgcaatgaaa	1740
caaaaattca ttgatcttag aaagagagca tttaaattcc caacaatggg taagaaagct	1800
agattaattt gtattccaac aacatcagga actggatcag aagttacacc atttgagtt	1860
atttcagatc atgaaacagg taagaaatat ccacttgctg attattcact tacaccatca	1920
gttgctattg ttgatecaat gtttactatg tcaacttcaa agagagctat tgetgatact	1980
ggacttgatg tattggttca tgcaacagaa gcatatgttt cagttatggc taatgaatat	2040
actgatggac ttgctagaga agcagttaaa ttagtctttg aaaatcttct taaatcatat	2100
aatggagatt tagaagcaag agaaaagatg cacaatgctg caacaattgc aggtatggca	2160
tttgcacag cattccttgg tatggacat tccatggcac ataaagtgg agcagcattc	2220
catcttccac atggtagatg tgtagcagta ttattaccac atgtcattag atataatgga	2280
caaaaacca gaaagcttgc aatgtggcca aaatataatt tctataaggc agaccaaaga	2340
tatatggaac ttgcacaaat ggttggactt aaatgtaata caccagctga aggagttgaa	2400
gcatttgcta aagcatgtga agaattaatg aaagccacag agactattac tggattcaag	2460
caagcaaata ttgatgaagc agcatggatg agtaaagtac cagaaatggc acttcttgca	2520
tttgaagatc aatgttcacc agctaatacca agagtcccaa tggttaagga tatggaaaag	2580
attctcaaag ctgcatatta tccaattgct tga	2613

<210> SEQ ID NO 49

<211> LENGTH: 2610

<212> TYPE: DNA

<213> ORGANISM: Artificial sequence

<220> FEATURE:

<223> OTHER INFORMATION: adh2 E. histolytica codon pair optimised

<400> SEQUENCE: 49

atgtccactc aacaaacat gaccgttgat gaacacatta accaattggt cagaaaggct	60
caagttgctt tgaaggaata cttgaaacca gaatacactc aagaaaagat cgattacatt	120
gtcaagaagg cttctgttgc tgctctagac caacactgtg ctttggctgc tgetgctgtc	180
gaagaaactg gtcgtggtat ctttgaagac aaagctacca agaacatttt cgcttgtaga	240
cacgtcactc acgaaatgag acacgccaag accgttggtg tcatcaacgt tgatccatta	300

-continued

tacgggtatca	ctgaaattgc	tgaaccagtc	gggtgtgtct	gtggtgtcac	cccagttacc	360
aacccaactt	ctactgccat	tttcaaactc	ttgatttcca	tcaagaccag	aaacccaatt	420
gtttttctct	tccacccttc	tgctttgaaa	tgttccatca	tggtctgcaa	gatcgtcaga	480
gatgctgcca	ttgctgctgg	tgctccagaa	aactgtatcc	aatggatcga	atttgggtgg	540
attgaagctt	ccaacaaatt	gatgaacat	cctgggtgtg	ctaccatctt	agctactgg	600
ggtaacgcta	tggtcaaggc	tgcttactct	tctggtaagc	cagctttggg	tgtcggtgct	660
ggtaacgtcc	caacttacat	cgaagagacc	tgtaatatca	agcaagctgc	taacgatgtt	720
gtcatgtcca	agtctttcga	caacgggatg	atctgtgcct	cgaacaagc	tgccatcatc	780
gacaaagaaa	tctacgacca	agttgttgaa	gaaatgaaga	ctttgggtgc	ttacttcac	840
aacgaagaag	aaaagggcaa	attggaaaaa	ttcatgttcg	gtgttaatgc	ttactctgct	900
gatgtcaaca	acgccagatt	gaacccaaag	tgtccaggta	tgtctccaca	atggttcgct	960
gaacaagtcg	gtatcaaggt	tccagaagac	tgtaacatca	tctgtgccgt	ttgtaaggaa	1020
gttgggtccaa	acgaaccatt	gaccagagaa	aagttgtctc	cagttttggc	cattttgaag	1080
gctgaaaaca	ctcaagatgg	tattgacaag	gctgaagcta	tggtcgaatt	caacggtcgt	1140
ggtcactctg	ctgccattca	ctccaatgac	aaggctgttg	ttgaaaaata	cgttttgacc	1200
atgaaggctt	gtcgtatctt	gcacaacact	ccatcttctc	aagggtgtat	cggttccatt	1260
tacaactaca	tctggccatc	tttacttta	ggttgtggtt	cttacgggtg	taactccgtt	1320
tctgccaatg	ttacctacca	caactgtgtg	aacatcaaga	gattggctga	cagaagaaac	1380
aacttacaat	ggttcagagt	cccaccaaag	atcttctctc	aacctcactc	cattagatac	1440
ttggctgaat	tgaaggaatt	gtccaagatt	ttcattgtct	ctgacagaat	gatgtacaaa	1500
ttgggttacg	ttgacagagt	tatggatgtc	ttgaagagaa	gatccaacga	agttgaaatt	1560
gaaatcttca	tcgatgttga	accagaccca	tccattcaaa	ccgtccaaaa	gggttttggt	1620
gtcatgaaca	ctttcgtgcc	agacaacatc	attgccattg	gtggtggttc	tgccatggat	1680
gctgccaaag	tcatgtggtt	attatacgaa	catccagaag	ctgatttctt	cgtatgaag	1740
caaaaattca	tcgatttaag	aaagagagct	ttcaagttcc	caacctggg	taagaaggcc	1800
agattaatct	gtatcccaac	cacttctggt	accggttctg	aagtcacccc	attcgctgtc	1860
atctctgacc	acgaaactgg	taagaagtat	ccattggctg	actactcttt	gaccccatcc	1920
gttgccattg	ttgacccaat	gtttaccatg	tccttgccct	agagagccat	tgttgacact	1980
ggtttggtatg	tcttagtcca	cgtactgaa	gcttacgttt	ctgttatggc	taacgaatac	2040
actgacggtt	tggccagaga	agctgtcaaa	ttggttttcg	aaaacttggt	gaaatcttac	2100
aacggtgact	tggaaagtcg	tgaagagatg	cacaacgctg	ctaccattgc	tggtatggcc	2160
tttgcttctg	ctttcttggt	tatggacat	tccatggctc	acaaggctcg	tgtgtctttc	2220
catttgccac	acggtagatg	tggtgocgtt	ttgttgcttc	acgttatcag	atacaacggt	2280
caaaagccaa	gaaagtggc	catgtggcca	aagtacaact	tctacaaggc	tgatcaaaga	2340
tacatggaat	tggctcaaat	ggtcgggttg	aagtgttaaca	ccccagctga	aggtgtcgaa	2400
gcctttgcca	aggcttgatg	agaattgatg	aaggctactg	aaaccatcac	tggtttcaag	2460
aaggccaaca	ttgatgaagc	tgcttggtatg	tccaaggttc	cagaaatggc	tctattggct	2520
ttcgaagacc	aatgttctcc	agctaaccac	agagtcccaa	tggttaagga	catggaaaag	2580

-continued

attttgaagg ctgcttacta cccaatcgct

2610

<210> SEQ ID NO 50

<211> LENGTH: 870

<212> TYPE: PRT

<213> ORGANISM: Entamoeba histolytica

<400> SEQUENCE: 50

```

Met Ser Thr Gln Gln Thr Met Thr Val Asp Glu His Ile Asn Gln Leu
1          5          10          15

Val Arg Lys Ala Gln Val Ala Leu Lys Glu Tyr Leu Lys Pro Glu Tyr
20        25        30

Thr Gln Glu Lys Ile Asp Tyr Ile Val Lys Lys Ala Ser Val Ala Ala
35        40        45

Leu Asp Gln His Cys Ala Leu Ala Ala Ala Val Glu Glu Thr Gly
50        55        60

Arg Gly Ile Phe Glu Asp Lys Ala Thr Lys Asn Ile Phe Ala Cys Glu
65        70        75        80

His Val Thr His Glu Met Arg His Ala Lys Thr Val Gly Ile Ile Asn
85        90        95

Val Asp Pro Leu Tyr Gly Ile Thr Glu Ile Ala Glu Pro Val Gly Val
100       105       110

Val Cys Gly Val Thr Pro Val Thr Asn Pro Thr Ser Thr Ala Ile Phe
115       120       125

Lys Ser Leu Ile Ser Ile Lys Thr Arg Asn Pro Ile Val Phe Ser Phe
130       135       140

His Pro Ser Ala Leu Lys Cys Ser Ile Met Ala Ala Lys Ile Val Arg
145       150       155       160

Asp Ala Ala Ile Ala Ala Gly Ala Pro Glu Asn Cys Ile Gln Trp Ile
165       170       175

Glu Phe Gly Gly Ile Glu Ala Ser Asn Lys Leu Met Asn His Pro Gly
180       185       190

Val Ala Thr Ile Leu Ala Thr Gly Gly Asn Ala Met Val Lys Ala Ala
195       200       205

Tyr Ser Ser Gly Lys Pro Ala Leu Gly Val Gly Ala Gly Asn Val Pro
210       215       220

Thr Tyr Ile Glu Lys Thr Cys Asn Ile Lys Gln Ala Ala Asn Asp Val
225       230       235       240

Val Met Ser Lys Ser Phe Asp Asn Gly Met Ile Cys Ala Ser Glu Gln
245       250       255

Ala Ala Ile Ile Asp Lys Glu Ile Tyr Asp Gln Val Val Glu Glu Met
260       265       270

Lys Thr Leu Gly Ala Tyr Phe Ile Asn Glu Glu Glu Lys Ala Lys Leu
275       280       285

Glu Lys Phe Met Phe Gly Val Asn Ala Tyr Ser Ala Asp Val Asn Asn
290       295       300

Ala Arg Leu Asn Pro Lys Cys Pro Gly Met Ser Pro Gln Trp Phe Ala
305       310       315       320

Glu Gln Val Gly Ile Lys Val Pro Glu Asp Cys Asn Ile Ile Cys Ala
325       330       335

Val Cys Lys Glu Val Gly Pro Asn Glu Pro Leu Thr Arg Glu Lys Leu
340       345       350

```

Ser	Pro	Val	Leu	Ala	Ile	Leu	Lys	Ala	Glu	Asn	Thr	Gln	Asp	Gly	Ile
		355				360						365			
Asp	Lys	Ala	Glu	Ala	Met	Val	Glu	Phe	Asn	Gly	Arg	Gly	His	Ser	Ala
		370				375				380					
Ala	Ile	His	Ser	Asn	Asp	Lys	Ala	Val	Val	Glu	Lys	Tyr	Ala	Leu	Thr
				390						395				400	
Met	Lys	Ala	Cys	Arg	Ile	Leu	His	Asn	Thr	Pro	Ser	Ser	Gln	Gly	Gly
				405				410						415	
Ile	Gly	Ser	Ile	Tyr	Asn	Tyr	Ile	Trp	Pro	Ser	Phe	Thr	Leu	Gly	Cys
		420						425				430			
Gly	Ser	Tyr	Gly	Gly	Asn	Ser	Val	Ser	Ala	Asn	Val	Thr	Tyr	His	Asn
		435				440						445			
Leu	Leu	Asn	Ile	Lys	Arg	Leu	Ala	Asp	Arg	Arg	Asn	Asn	Leu	Gln	Trp
		450				455				460					
Phe	Arg	Val	Pro	Pro	Lys	Ile	Phe	Phe	Glu	Pro	His	Ser	Ile	Arg	Tyr
				470						475				480	
Leu	Ala	Glu	Leu	Lys	Glu	Leu	Ser	Lys	Ile	Phe	Ile	Val	Ser	Asp	Arg
				485				490						495	
Met	Met	Tyr	Lys	Leu	Gly	Tyr	Val	Asp	Arg	Val	Met	Asp	Val	Leu	Lys
		500						505				510			
Arg	Arg	Ser	Asn	Glu	Val	Glu	Ile	Glu	Ile	Phe	Ile	Asp	Val	Glu	Pro
		515				520						525			
Asp	Pro	Ser	Ile	Gln	Thr	Val	Gln	Lys	Gly	Leu	Ala	Val	Met	Asn	Thr
		530				535				540					
Phe	Gly	Pro	Asp	Asn	Ile	Ile	Ala	Ile	Gly	Gly	Gly	Ser	Ala	Met	Asp
				550						555				560	
Ala	Ala	Lys	Ile	Met	Trp	Leu	Leu	Tyr	Glu	His	Pro	Glu	Ala	Asp	Phe
				565				570						575	
Phe	Ala	Met	Lys	Gln	Lys	Phe	Ile	Asp	Leu	Arg	Lys	Arg	Ala	Phe	Lys
		580						585				590			
Phe	Pro	Thr	Met	Gly	Lys	Lys	Ala	Arg	Leu	Ile	Cys	Ile	Pro	Thr	Thr
		595				600						605			
Ser	Gly	Thr	Gly	Ser	Glu	Val	Thr	Pro	Phe	Ala	Val	Ile	Ser	Asp	His
		610				615				620					
Glu	Thr	Gly	Lys	Lys	Tyr	Pro	Leu	Ala	Asp	Tyr	Ser	Leu	Thr	Pro	Ser
				630						635				640	
Val	Ala	Ile	Val	Asp	Pro	Met	Phe	Thr	Met	Ser	Leu	Pro	Lys	Arg	Ala
				645				650						655	
Ile	Ala	Asp	Thr	Gly	Leu	Asp	Val	Leu	Val	His	Ala	Thr	Glu	Ala	Tyr
		660						665				670			
Val	Ser	Val	Met	Ala	Asn	Glu	Tyr	Thr	Asp	Gly	Leu	Ala	Arg	Glu	Ala
		675				680						685			
Val	Lys	Leu	Val	Phe	Glu	Asn	Leu	Leu	Lys	Ser	Tyr	Asn	Gly	Asp	Leu
		690				695				700					
Glu	Ala	Arg	Glu	Lys	Met	His	Asn	Ala	Ala	Thr	Ile	Ala	Gly	Met	Ala
				710						715				720	
Phe	Ala	Ser	Ala	Phe</											

-continued

755	760	765
Trp Pro Lys Tyr Asn Phe	Tyr Lys Ala Asp Gln Arg	Tyr Met Glu Leu
770	775	780
Ala Gln Met Val Gly Leu	Lys Cys Asn Thr Pro	Ala Glu Gly Val Glu
785	790	800
Ala Phe Ala Lys Ala Cys	Glu Glu Leu Met Lys	Ala Thr Glu Thr Ile
805	810	815
Thr Gly Phe Lys Lys Ala	Asn Ile Asp Glu Ala	Ala Trp Met Ser Lys
820	825	830
Val Pro Glu Met Ala Leu	Leu Ala Phe Glu Asp	Gln Cys Ser Pro Ala
835	840	845
Asn Pro Arg Val Pro Met	Val Lys Asp Met Glu	Lys Ile Leu Lys Ala
850	855	860
Ala Tyr Tyr Pro Ile Ala		
865	870	

<210> SEQ ID NO 51

<211> LENGTH: 2658

<212> TYPE: DNA

<213> ORGANISM: *Piromyces* sp. E2

<400> SEQUENCE: 51

```

atgtccggat tacaaatggt ccaaaacctt tctctttacg gtagtctcgc cgaaatcgat      60
actagcgaaa agcttaacga agctatggac aaattaactg ctgcccaaga acaattcaga      120
gaatacaacc aagaacaagt tgacaaaatc ttcaaggctg ttgctttagc tgcttctcaa      180
aaccgtgttg ctttcgctaa gtacgcacac gaagaaaccc aaaaggggtg tttcgaagat      240
aagggttatca agaacgaatt cgctgctgat tacatttacc acaagtactg caatgacaag      300
accgccggta tcattgaata tgatgaagcc aatggtctta tggaaattgc tgaaccagtt      360
ggtcaggttg ttggtattgc tccagttact aacccaactt ctactatcat ctacaagtct      420
ttaattgcct taaagaccgg taactgtatt atcttctcac cacatccagg agctcacaag      480
gcctctgttt tcggtgttaa ggtcttacac caagctgctg ttaaggctgg tgccccagaa      540
aactgtattc aatatcatct cccaagatg gatttaacta ctgaattatt acaccaccaa      600
aagactcggt tcatttgggc tactgggtgg ccagggttag ttacgcctc ttacacttct      660
ggtaagccag ctcttggtgg tgggccaggt aatgctccag ctcttattga tgaaacttgt      720
gatatgaacg aagctgttgg ttctatcggt gtttctaaga ctttcgattg tggatatgatc      780
tgtgccactg aaaacgctgt tgcgttgtgc gaatctgtct acgaaaactt cgttgctacc      840
atgaagaagc gtggtgccta cttcatgact ccagaagaaa ccaagaaggc ttctaacctt      900
cttttcggag aaggatatgag attaaatgct aaggctgttg gtcaaactgc caagacttta      960
gctgaaatgg ccggttttca agtcccagaa aacaccgttg ttctctgtgg tgaagcttct      1020
gaagttaaat tcgaagaacc aatggctcac gaaaagttaa ctactatcct cggtatctac      1080
aaggctaagg actttgacga tgggtgcaga ttatgtaagg aattagttac tttcgggtgg      1140
aagggtcaca ctgctgttct ctacaccaac caaaacaaca aggaccgtat tgaaaagtac      1200
caaaacgaag ttccagcctt ccacatctta gttgacatgc catcttcctt cggttgtatt      1260
ggtgatatgt acaacttcgg tcttgctcca gctcttacca ttacttgtgg tactatgggt      1320
ggtggttctt cctctgataa cattgggtcca aagcacttac ttaacatcaa gcgtgttggt      1380

```

-continued

```

atgagacgcg aaaacatgct ttggttcaag attccaaagt ctgtctactt caagcgtgct 1440
atcctttctg aagctttatc tgacttacgt gacaccacaca agcgtgctat cattattacc 1500
gatagaacta tgactatggt aggtcaaact gacaagatca ttaaggcttg tgaaggatcat 1560
ggtatggtct gcactgtcta cgataagggt gtcccagatc caactatcaa gtgtattatg 1620
gaagggtgta atgaaatgaa cgtcttcaag ccagatttag ctattgctct tgggtggtgt 1680
tctgctatgg atgccgctaa gatgatgcgt ttattctacg aataccacaga ccaagactta 1740
caagatattg ctactcggtt cgtcgatata cgtaagcgtg ttgttggttg tccaaagctt 1800
ggtagactta ttaagactct tgtctgtatc ccaactacct ctggtactgg tgcgaaggt 1860
actccattcg ctgctgttac ctctgaagaa ggctgtaagt acccattagt cgactacgaa 1920
cttactccag atatggctat tgttgatcca gaattcgtg ttggtatgcc aaagcgttta 1980
acttcttgga ctggtattga tgcctttacc cagccattg aatcttacgt ttctattatg 2040
gctactgact tcaactagacc atactctctc cgtgctgttg gtcttatctt cgaatccctt 2100
tcccttgctt acaacaacgg taaggatatt gaagctcgtg aaaagatgca caatgcttct 2160
gctattgctg gtatggcctt tgccaacgct ttccttggtt gttgtcactc tgttgcacac 2220
caacttggtt ccgtctacca cattccacac ggtcttgcca acgctttaat gctttctcac 2280
atcattaagt acaacgctac tgactctcca gttaagatgg gtaccttccc acaatacaag 2340
taccacaag ctatgcgtca ctacgtgaa attgctgaac tcttattacc accaactcaa 2400
gttgtaaga tgactgatgt tgataagggt caatacttaa ttgaccgtgt tgaacaatta 2460
aaggctgacg ttggtattcc aaagtctatt aaggaaactg gaatgggtac tgaagaagac 2520
ttcttcaaca aggttgacca agttgctatc atggccttcg atgaccaatg tactggtgct 2580
aaccacggtt acccattagt ttctgaatta aaacaattaa tgattgatgc ctggaacggt 2640
gttggtccaa agctctaa 2658

```

<210> SEQ ID NO 52

<211> LENGTH: 885

<212> TYPE: PRT

<213> ORGANISM: *Piromyces* sp. E2

<400> SEQUENCE: 52

```

Met Ser Gly Leu Gln Met Phe Gln Asn Leu Ser Leu Tyr Gly Ser Leu
1           5           10          15
Ala Glu Ile Asp Thr Ser Glu Lys Leu Asn Glu Ala Met Asp Lys Leu
20          25          30
Thr Ala Ala Gln Glu Gln Phe Arg Glu Tyr Asn Gln Glu Gln Val Asp
35          40          45
Lys Ile Phe Lys Ala Val Ala Leu Ala Ala Ser Gln Asn Arg Val Ala
50          55          60
Phe Ala Lys Tyr Ala His Glu Glu Thr Gln Lys Gly Val Phe Glu Asp
65          70          75          80
Lys Val Ile Lys Asn Glu Phe Ala Ala Asp Tyr Ile Tyr His Lys Tyr
85          90          95
Cys Asn Asp Lys Thr Ala Gly Ile Ile Glu Tyr Asp Glu Ala Asn Gly
100         105         110
Leu Met Glu Ile Ala Glu Pro Val Gly Pro Val Val Gly Ile Ala Pro
115        120        125

```

-continued

Val	Thr	Asn	Pro	Thr	Ser	Thr	Ile	Ile	Tyr	Lys	Ser	Leu	Ile	Ala	Leu
130						135					140				
Lys	Thr	Arg	Asn	Cys	Ile	Ile	Phe	Ser	Pro	His	Pro	Gly	Ala	His	Lys
145					150					155					160
Ala	Ser	Val	Phe	Val	Val	Lys	Val	Leu	His	Gln	Ala	Ala	Val	Lys	Ala
				165					170					175	
Gly	Ala	Pro	Glu	Asn	Cys	Ile	Gln	Ile	Ile	Phe	Pro	Lys	Met	Asp	Leu
			180					185					190		
Thr	Thr	Glu	Leu	Leu	His	His	Gln	Lys	Thr	Arg	Phe	Ile	Trp	Ala	Thr
		195					200					205			
Gly	Gly	Pro	Gly	Leu	Val	His	Ala	Ser	Tyr	Thr	Ser	Gly	Lys	Pro	Ala
	210					215					220				
Leu	Gly	Gly	Gly	Pro	Gly	Asn	Ala	Pro	Ala	Leu	Ile	Asp	Glu	Thr	Cys
225					230					235					240
Asp	Met	Asn	Glu	Ala	Val	Gly	Ser	Ile	Val	Val	Ser	Lys	Thr	Phe	Asp
			245						250					255	
Cys	Gly	Met	Ile	Cys	Ala	Thr	Glu	Asn	Ala	Val	Val	Val	Val	Glu	Ser
		260						265					270		
Val	Tyr	Glu	Asn	Phe	Val	Ala	Thr	Met	Lys	Lys	Arg	Gly	Ala	Tyr	Phe
	275						280					285			
Met	Thr	Pro	Glu	Glu	Thr	Lys	Lys	Ala	Ser	Asn	Leu	Leu	Phe	Gly	Glu
	290					295					300				
Gly	Met	Arg	Leu	Asn	Ala	Lys	Ala	Val	Gly	Gln	Thr	Ala	Lys	Thr	Leu
305					310					315					320
Ala	Glu	Met	Ala	Gly	Phe	Glu	Val	Pro	Glu	Asn	Thr	Val	Val	Leu	Cys
			325						330					335	
Gly	Glu	Ala	Ser	Glu	Val	Lys	Phe	Glu	Glu	Pro	Met	Ala	His	Glu	Lys
		340					345						350		
Leu	Thr	Thr	Ile	Leu	Gly	Ile	Tyr	Lys	Ala	Lys	Asp	Phe	Asp	Asp	Gly
	355					360						365			
Val	Arg	Leu	Cys	Lys	Glu	Leu	Val	Thr	Phe	Gly	Gly	Lys	Gly	His	Thr
	370					375					380				
Ala	Val	Leu	Tyr	Thr	Asn	Gln	Asn	Asn	Lys	Asp	Arg	Ile	Glu	Lys	Tyr
385					390					395					400
Gln	Asn	Glu	Val	Pro	Ala	Phe	His	Ile	Leu	Val	Asp	Met	Pro	Ser	Ser
			405						410					415	
Leu	Gly	Cys	Ile	Gly	Asp	Met	Tyr	Asn	Phe	Arg	Leu	Ala	Pro	Ala	Leu
		420					425					430			
Thr	Ile	Thr	Cys	Gly	Thr	Met	Gly	Gly	Gly	Ser	Ser	Ser	Asp	Asn	Ile
	435					440						445			
Gly	Pro	Lys	His	Leu	Leu	Asn	Ile	Lys	Arg	Val	Gly	Met	Arg	Arg	Glu
	450					455					460				
Asn	Met	Leu	Trp	Phe	Lys	Ile	Pro	Lys	Ser	Val	Tyr	Phe	Lys	Arg	Ala
465					470					475					480
Ile	Leu	Ser	Glu	Ala	Leu	Ser	Asp	Leu	Arg	Asp	Thr	His	Lys	Arg	Ala
			485						490					495	
Ile	Ile	Ile	Thr	Asp	Arg	Thr	Met	Thr	Met	Leu	Gly	Gln	Thr	Asp	Lys
			500					505					510		
Ile	Ile	Lys	Ala	Cys	Glu	Gly	His	Gly	Met	Val	Cys	Thr	Val	Tyr	Asp
	515						520					525			

Lys	Val	Val	Pro	Asp	Pro	Thr	Ile	Lys	Cys	Ile	Met	Glu	Gly	Val	Asn
530															
Glu	Met	Asn	Val	Phe	Lys	Pro	Asp	Leu	Ala	Ile	Ala	Leu	Gly	Gly	Gly
545															
Ser	Ala	Met	Asp	Ala	Ala	Lys	Met	Met	Arg	Leu	Phe	Tyr	Glu	Tyr	Pro
565															
Asp	Gln	Asp	Leu	Gln	Asp	Ile	Ala	Thr	Arg	Phe	Val	Asp	Ile	Arg	Lys
580															
Arg	Val	Val	Gly	Cys	Pro	Lys	Leu	Gly	Arg	Leu	Ile	Lys	Thr	Leu	Val
595															
Cys	Ile	Pro	Thr	Thr	Ser	Gly	Thr	Gly	Ala	Glu	Val	Thr	Pro	Phe	Ala
610															
Val	Val	Thr	Ser	Glu	Glu	Gly	Arg	Lys	Tyr	Pro	Leu	Val	Asp	Tyr	Glu
625															
Leu	Thr	Pro	Asp	Met	Ala	Ile	Val	Asp	Pro	Glu	Phe	Ala	Val	Gly	Met
645															
Pro	Lys	Arg	Leu	Thr	Ser	Trp	Thr	Gly	Ile	Asp	Ala	Leu	Thr	His	Ala
660															
Ile	Glu	Ser	Tyr	Val	Ser	Ile	Met	Ala	Thr	Asp	Phe	Thr	Arg	Pro	Tyr
675															
Ser	Leu	Arg	Ala	Val	Gly	Leu	Ile	Phe	Glu	Ser	Leu	Ser	Leu	Ala	Tyr
690															
Asn	Asn	Gly	Lys	Asp	Ile	Glu	Ala	Arg	Glu	Lys	Met	His	Asn	Ala	Ser
705															
Ala	Ile	Ala	Gly	Met	Ala	Phe	Ala	Asn	Ala	Phe	Leu	Gly	Cys	Cys	His
725															
Ser	Val	Ala	His	Gln	Leu	Gly	Ser	Val	Tyr	His	Ile	Pro	His	Gly	Leu
740															
Ala	Asn	Ala	Leu	Met	Leu	Ser	His	Ile	Ile	Lys	Tyr	Asn	Ala	Thr	Asp
755															
Ser	Pro	Val	Lys	Met	Gly	Thr	Phe	Pro	Gln	Tyr	Lys	Tyr	Pro	Gln	Ala
770															
Met	Arg	His	Tyr	Ala	Glu	Ile	Ala	Glu	Leu	Leu	Leu	Pro	Pro	Thr	Gln
785															
Val	Val	Lys	Met	Thr	Asp	Val	Asp	Lys	Val	Gln	Tyr	Leu	Ile	Asp	Arg
805															
Val	Glu	Gln	Leu	Lys	Ala	Asp	Val	Gly	Ile	Pro	Lys	Ser	Ile	Lys	Glu
820															
Thr	Gly	Met	Val	Thr	Glu	Glu	Asp	Phe	Phe	Asn	Lys	Val	Asp	Gln	Val
835															
Ala	Ile	Met	Ala	Phe	Asp	Asp	Gln	Cys	Thr	Gly	Ala	Asn	Pro	Arg	Tyr
850															
Pro	Leu	Val	Ser	Glu	Leu	Lys	Gln	Leu	Met	Ile	Asp	Ala	Trp	Asn	Gly
865															
Val	Val	Pro	Lys	Leu											
885															

1. A method of identifying a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) a yeast cell comprising:

providing a mutated yeast cell, wherein said mutation comprises an inactivation of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS);

transforming said mutated yeast cell with an expression vector comprising at least one heterologous nucleotide sequence operably linked to a promoter functional in yeast and said heterologous nucleotide sequence encoding a candidate polypeptide having potential enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl CoA;

testing said recombinant mutated yeast cell for its ability to grow on minimal medium containing glucose as sole carbon source, and

identifying said candidate polypeptide as a heterologous polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl CoA in (the cytosol of) said yeast cell when growth of said cell is observed.

2. Method according to claim 1, wherein said yeast cell is a cell of *Saccharomyces cerevisiae* and wherein said heterologous nucleotide sequence is codon pair optimized for expression in *Saccharomyces cerevisiae*.

3. Method according to claim 2, wherein said mutation comprises an inactivation of the gene for acetyl-CoA synthetase isoform 2 (acs2).

4. Method according to claim 1, wherein said candidate polypeptide having enzymatic activity for converting acetaldehyde into acetyl-CoA is a (putative) acetylating acetaldehyde dehydrogenase (acdh).

5. A vector for the expression of heterologous polypeptides in yeast, said vector comprising a heterologous nucleotide sequence operably linked to a promoter functional in yeast and said heterologous nucleotide sequence encoding a polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell.

6. Vector according to claim 5, wherein said polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA.

7. Vector according to claim 5, wherein said polypeptide has more than 50%, preferably more than 60%, 70%, 80%,

90%, or 95% sequence identity with the amino acid sequence selected from SEQ ID NO: 19, 22, 25, 28 and 52.

8. Vector according to claim 5 for expression in *Saccharomyces cerevisiae*, wherein said heterologous nucleotide sequence is codon pair optimized for expression in *Saccharomyces cerevisiae*.

9. Expression vector according to claim 8, wherein said heterologous nucleotide sequence is selected from SEQ ID NO: 20, 23, 26, 29 and 51.

10. A recombinant yeast cell comprising a vector of claim 5.

11. A recombinant yeast cell comprising a heterologous nucleotide sequence encoding a polypeptide having enzymatic activity for converting pyruvate, acetaldehyde or acetate into acetyl-CoA in (the cytosol of) said yeast cell.

12. Yeast cell according to claim 10, further comprising an inactivation of at least one gene of the (PDH) by-pass, selected from the genes encoding the enzymes pyruvate decarboxylase (PDC), acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS).

13. Yeast cell according to claim 10, wherein the yeast cell comprises an inactivation of a gene encoding an acetyl-CoA synthase.

14. Yeast cell according to claim 10, wherein said cell shows growth on minimal medium containing glucose as sole carbon source.

15. Yeast cell according to claim 10, further comprising an inactivation of a gene encoding an enzyme that catalyses the conversion of acetaldehyde into ethanol, preferably an alcohol dehydrogenase.

16. Yeast cell according to claim 10, further comprising one or more introduced genes encoding a recombinant pathway for the formation of 1-butanol from acetyl-CoA.

17. Yeast cell according to claim 16, wherein said one or more introduced genes encode enzymes that produce acetoacetyl-CoA, 3-hydroxybutyryl-CoA, crotonyl-CoA, butyryl-CoA, butylaldehyde and/or 1-butanol.

18. Yeast cell according to claim 10, wherein said yeast is *Saccharomyces cerevisiae*.

19. A method of producing a fermentation product, comprising the steps of fermenting a suitable carbon substrate with a yeast cell according to claim 10 and recovering the fermentation product produced during said fermentation.

20. Method according to claim 19, wherein the fermentation product is butanol.

* * * * *