



(86) Date de dépôt PCT/PCT Filing Date: 2011/11/18
(87) Date publication PCT/PCT Publication Date: 2012/05/24
(45) Date de délivrance/Issue Date: 2019/03/26
(85) Entrée phase nationale/National Entry: 2013/10/22
(86) N° demande PCT/PCT Application No.: NL 2011/050787
(87) N° publication PCT/PCT Publication No.: 2012/067510
(30) Priorités/Priorities: 2010/11/18 (US61/415,054);
2010/11/18 (EP10191736.7); 2011/04/05 (US61/471,836)

(51) Cl.Int./Int.Cl. *C12P 7/06* (2006.01),
C12N 1/32 (2006.01), *C12N 9/02* (2006.01),
C12N 9/04 (2006.01), *C12N 9/10* (2006.01),
C12N 9/12 (2006.01), *C12P 7/40* (2006.01)
(72) Inventeurs/Inventors:
DE BONT, JOHANNES ADRIANUS MARIA, NL;
TEUNISSEN, ALOYSIUS WILHELMUS RUDOLPHUS
HUBERTUS, NL
(73) Propriétaire/Owner:
DSM IP ASSETS B.V., NL
(74) Agent: SMART & BIGGAR

(54) Titre : SOUCHES DE LEVURES MODIFIEES POUR PRODUIRE DE L'ETHANOL A PARTIR DE GLYCEROL
(54) Title: YEAST STRAINS ENGINEERED TO PRODUCE ETHANOL FROM GLYCEROL

(57) **Abrégé/Abstract:**

The present invention relates to genetically modified yeast cells comprising exogenous genes coding for pyruvate formate lyase and acetaldehyde dehydrogenase activities. The yeast cells further comprises genetic modifications that improve glycerol utilization such as modifications that increases NAD⁺-linked glycerol dehydrogenase activity, and preferably one or more of dihydroxyacetone kinase activity and transport of glycerol into the cell. The yeast cell further preferably comprises a functional exogenous xylose isomerase gene and/or functional exogenous genes which confer to the cell the ability to convert L-arabinose into D-xylulose 5-phosphate and they may comprise a genetic modification that increase acetyl-CoA synthetase activity. The process is further characterised in that glycerol is present in or fed into the culture medium, whereby the modified yeast cell ferments the hexoses, pentoses, acetic acid and glycerol to ethanol. The invention further relates to processes for producing a fermentation product and formate from carbon sources comprising glycerol and at least one of a hexose and a pentose, wherein the genetically modified yeast cells are used to ferment the carbon sources to the fermentation product and formate. A preferred fermentation product is ethanol.

- (51) **International Patent Classification:**

<i>C12P 7/06</i> (2006.01)	<i>C12N 9/04</i> (2006.01)
<i>C12N 1/32</i> (2006.01)	<i>C12N 9/12</i> (2006.01)
<i>C12N 9/10</i> (2006.01)	<i>C12P 7/40</i> (2006.01)
<i>C12N 9/02</i> (2006.01)	

(21) **International Application Number:** PCT/NL2011/050787

(22) **International Filing Date:** 18 November 2011 (18.11.2011)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

10191736.7	18 November 2010 (18.11.2010)	EP
61/415,054	18 November 2010 (18.11.2010)	US
61/471,836	5 April 2011 (05.04.2011)	US

(71) **Applicant** (*for all designated States except US*): **C5 YEAST COMPANY B.V.** [NL/NL]; Lelyweg 29, NL-4612 PS Bergen op Zoom (NL).

(72) **Inventors; and**

(75) **Inventors/Applicants** (*for US only*): **DE BONT, Johannes Adrianus Maria** [NL/NL]; Lawickse Allee 68, NL-6707 AK Wageningen (NL). **TEUNISSEN, Aloysius Wilhelmus Rudolphus Hubertus** [NL/NL]; Tjaskerlaan 28, NL-3052 HP Rotterdam (NL).

(74) **Agent:** **HUTTER, Hans**; J.W. Frisolaan 13, NL-2517 JS The Hague (NL).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

 - *with international search report (Art. 21(3))*
 - *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
 - *with sequence listing part of description (Rule 5.2(a))*

- (54) **Title:** Yeast strains engineered to produce ethanol from glycerol
- (57) **Abstract:** The present invention relates to genetically modified yeast cells comprising exogenous genes coding for pyruvate formate lyase and acetaldehyde dehydrogenase activities. The yeast cells further comprises genetic modifications that improve glycerol utilization such as modifications that increases NAD⁺ -linked glycerol dehydrogenase activity, and preferably one or more of dihydroxyacetone kinase activity and transport of glycerol into the cell. The yeast cell further preferably comprises a functional exogenous xylose isomerase gene and/or functional exogenous genes which confer to the cell the ability to convert L-arabinose into D-xylulose 5-phosphate and they may comprise a genetic modification that increase acetyl-CoA synthetase activity. The process is further characterised in that glycerol is present in or fed into the culture medium, whereby the modified yeast cell ferments the hexoses, pentoses, acetic acid and glycerol to ethanol. The invention further relates to processes for producing a fermentation product and formate from carbon sources comprising glycerol and at least one of a hexose and a pentose, wherein the genetically modified yeast cells are used to ferment the carbon sources to the fermentation product and formate. A preferred fermentation product is ethanol.

Yeast strains engineered to produce ethanol from glycerolField of the invention

The present invention relates to metabolic engineering in microorganisms such as yeast. In particular the invention relates to yeast strains that have been engineered to produce ethanol from glycerol. These strains have retained their natural ability to produce ethanol from hexoses (glucose, fructose, galactose, etc) and comprise an engineered ability to produce ethanol from pentoses like xylose. The invention further relates to the processes wherein the engineered strains of the invention produce ethanol from glycerol, either as main fermentation feedstock, or concomitantly with one or more of hexoses and pentoses.

Background of the invention

Both the formation and the degradation of glycerol in the yeast *Saccharomyces cerevisiae* are important processes that have been studied for more than a century. During the last decades, the emphasis has been on (1) suppression of the formation of glycerol during ethanol production. In the beginning of the previous century, (2) the optimization of glycerol production from sugars has been studied. Both of these aspects have profound economic impacts. More recently, a fully new aspect has been introduced in this glycerol field. It deals with (3) the anaerobic fermentation of glycerol into ethanol.

(1) Suppression of glycerol formation during ethanol production.

Bioethanol is produced by *Saccharomyces cerevisiae* from a range of substrates including lignocellulosic hydrolysates of non-food feedstocks (e.g. energy crops and agricultural residues). One problem of yeast-based ethanol production is that during anaerobic ethanolic fermentation of sugar feedstocks, substantial amounts of glycerol are invariably formed as a by-product. Sugar dissimilation during anaerobic growth of *S. cerevisiae* occurs via alcoholic fermentation. In this process, the NADH formed in glycolysis is reoxidised via a NAD^+ dependent alcohol dehydrogenase which converts acetaldehyde (formed by decarboxylation of pyruvate) to ethanol. This dissimilatory pathway is redox-neutral and can therefore not compensate a net reduction of NAD^+ to NADH occurring elsewhere in metabolism. Such net reduction of NAD^+ to NADH occurs in assimilation when yeast biomass is synthesized under anaerobic conditions

from sugars and e.g. ammonia. Under anaerobic conditions, NADH reoxidation in *S. cerevisiae* to compensate for the assimilatory-associated NADH-formation is mainly dependent on reduction of at least part of the sugar carbon source to glycerol, resulting in a lower ethanol yield. To address this problem, several metabolic engineering
5 approaches have been explored to reduce or eliminate glycerol production in anaerobically grown *S. cerevisiae*. Two examples are:

Nissen et al (2000) disclose an approach in which the enzyme NADPH-dependent glutamate dehydrogenase was deleted. Glutamine synthetase and glutamate synthase were overexpressed. The resulting strain no longer synthesized glutamate
10 from ammonium and 2-oxoglutarate via a NADPH-requiring route, but rather via a NADH- and ATP-requiring pathway. The resulting strain had a 10% higher yield in ethanol and a 38% lower glycerol yield than the wild type strain.

Guadalupe Medina et al. (2009, Appl. Environ. Microbiol., 76: 190–195) disclose a *S. cerevisiae* strain wherein production of the by-product glycerol is eliminated by
15 disruption of the endogenous NAD-dependent glycerol 3-phosphate dehydrogenase genes (*GPD1* and *GPD2*). Expression of the *E. coli mhpF* gene, encoding the acetylating NAD-dependent acetaldehyde dehydrogenase restored the ability of the GPD-disrupted strain to grow anaerobically. However, the GPD-disrupted strain could only grow anaerobically if the medium is supplemented with acetic acid.

20 (2) Optimization of glycerol production from sugars.

Production of glycerol from sugars, as opposed to minimizing its production during ethanol formation, has been an economically important process during World War 1 in Germany. In recent years, a renewed interest in this process is seen, as witnessed for instance by the work of Overkamp et al (2002, Appl. Environ. Microbiol.
25 68:2814-21).

(3) Fermentation of glycerol into ethanol.

The fermentation of glycerol into ethanol is not feasible in normal *S. cerevisiae* and under fully anoxic cultivation conditions due to an unbalanced balance for NADH. Glycerol is more reduced than ethanol and hence the organism cannot dispose of its
30 excess NADH by reoxidation under normal anoxic conditions. Some initial work has been done in the field of fermentation of glycerol into ethanol: Yu et al. (2010, Bioresour. Technol. 101(11):4157-61. Epub 2010 Feb 9) disclose *S. cerevisiae* strains metabolically engineered for improved ethanol production from glycerol by

simultaneous overexpression of glycerol dehydrogenase (*GCH*), dihydroxyacetone kinase (*DAK*) and the glycerol uptake protein (*GUP1*). In a later communication, Yu et al (2010, J. Biotechnol. doi:10.1016/j.jbiotec.2010.09.932) disclose an optimisation of their strain for ethanol production from glycerol by deleting two glycerol production
5 genes, *FPS1* and *GPD2*. It was shown that ethanol production from glycerol is possible in their strains. However, the increase in yield was dependent on micro aerobic conditions, which can be explained by the requirement for NADH oxidation. The excess of NADH produced in the path from glycerol to ethanol apparently could only be reoxidised via an oxygen-dependent reaction.

10 Waks and Silver (2009, Appl. Environ. Microbiol., 75:1867–1875) disclose a synthetic dual-organism system for biohydrogen production. In a first step formate is produced by an engineered *S. cerevisiae* strain wherein a formate-overproducing pathway has been implemented. In a second step the formate produced by the engineered yeast is processed into hydrogen by *Escherichia coli*. The *S. cerevisiae*
15 strain was engineered to produce formate by expressing the anaerobic enzyme pyruvate formate lyase (PFL) from *E. coli*. Formate production was further increased by also introducing expression of a downstream enzyme, the AdhE of *E. coli*, the bifunctional enzyme that reduces acetyl-CoA generated by PFL into ethanol.

It is an object of the present invention to provide for yeasts that are capable of
20 producing ethanol from glycerol while retaining their abilities of fermenting hexoses (glucose, fructose, galactose, etc) as well as pentoses like xylose, as well as processes wherein these strains are used for the production of ethanol and/or other fermentation products.

25 Description of the invention

Definitions

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. In the art, "identity" also means
30 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. "Similarity" between two amino acid sequences is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the

sequence of a second polypeptide. "Identity" and "similarity" can be readily calculated by known methods. The terms "sequence identity" or "sequence similarity" means that two (poly)peptide or two nucleotide sequences, when optimally aligned, preferably over the entire length (of at least the shortest sequence in the comparison) and maximizing the number of matches and minimizes the number of gaps such as by the programs ClustalW (1.83), GAP or BESTFIT using default parameters, share at least a certain percentage of sequence identity as defined elsewhere herein. GAP uses the Needleman and Wunsch global alignment algorithm to align two sequences over their entire length, maximizing the number of matches and minimizes the number of gaps. Generally, the GAP default parameters are used, with a gap creation penalty = 50 (nucleotides) / 8 (proteins) and gap extension penalty = 3 (nucleotides) / 2 (proteins). For nucleotides the default scoring matrix used is nws gapdna and for proteins the default scoring matrix is Blosom62 (Henikoff & Henikoff, 1992, PNAS 89, 915-919). A preferred multiple alignment program for aligning protein sequences of the invention is ClustalW (1.83) using a blosom matrix and default settings (Gap opening penalty:10; Gap extension penalty: 0.05). It is clear than when RNA sequences are said to be essentially similar or have a certain degree of sequence identity with DNA sequences, thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. Sequence alignments and scores for percentage sequence identity may be determined using computer programs, such as the GCG Wisconsin Package, Version 10.3, available from Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA or the open-source software Emboss for Windows (current version 2.10.0-0.8). Alternatively percent similarity or identity may be determined by searching against databases such as FASTA, BLAST, etc.

A variant of a nucleotide or amino acid sequence disclosed herein may also be defined as a nucleotide or amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the nucleotide or amino acid sequence specifically disclosed herein (e.g. in de the sequence listing).

Optionally, in determining the degree of amino acid similarity, the skilled person may also take into account so-called "conservative" amino acid substitutions, as will be clear to the skilled person. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and

isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulphur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine. Substitutional variants of the amino acid sequence disclosed herein are those in which at least one residue in the disclosed sequences has been removed and a different residue inserted in its place. Preferably, the amino acid change is conservative. Preferred conservative substitutions for each of the naturally occurring amino acids are as follows: Ala to ser; Arg to lys; Asn to gln or his; Asp to glu; Cys to ser or ala; Gln to asn; Glu to asp; Gly to pro; His to asn or gln; Ile to leu or val; Leu to ile or val; Lys to arg; gln or glu; Met to leu or ile; Phe to met, leu or tyr; Ser to thr; Thr to ser; Trp to tyr; Tyr to trp or phe; and, Val to ile or leu.

Nucleotide sequences of the invention may also be defined by their capability to hybridise with parts of specific nucleotide sequences disclosed herein, respectively, under moderate, or preferably under stringent hybridisation conditions. Stringent hybridisation conditions are herein defined as conditions that allow a nucleic acid sequence of at least about 25, preferably about 50 nucleotides, 75 or 100 and most preferably of about 200 or more nucleotides, to hybridise at a temperature of about 65°C in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at 65°C in a solution comprising about 0.1 M salt, or less, preferably 0.2 x SSC or any other solution having a comparable ionic strength. Preferably, the hybridisation is performed overnight, i.e. at least for 10 hours and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having about 90% or more sequence identity.

Moderate conditions are herein defined as conditions that allow a nucleic acid sequences of at least 50 nucleotides, preferably of about 200 or more nucleotides, to hybridise at a temperature of about 45°C in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at room temperature in a solution comprising about 1 M salt, preferably 6 x

SSC or any other solution having a comparable ionic strength. Preferably, the hybridisation is performed overnight, i.e. at least for 10 hours, and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having up to 50% sequence identity. The person skilled in the art will be able to modify these hybridisation conditions in order to specifically identify sequences varying in identity between 50% and 90%.

A “nucleic acid construct” or “nucleic acid vector” is herein understood to mean a man-made nucleic acid molecule resulting from the use of recombinant DNA technology. The term “nucleic acid construct” therefore does not include naturally occurring nucleic acid molecules although a nucleic acid construct may comprise (parts of) naturally occurring nucleic acid molecules. The terms “expression vector” or “expression construct” refer to nucleotide sequences that are capable of affecting expression of a gene in host cells or host organisms compatible with such sequences. These expression vectors typically include at least suitable transcription regulatory sequences and optionally, 3’ transcription termination signals. Additional factors necessary or helpful in effecting expression may also be present, such as expression enhancer elements. The expression vector will be introduced into a suitable host cell and be able to effect expression of the coding sequence in an in vitro cell culture of the host cell. The expression vector will be suitable for replication in the host cell or organism of the invention.

As used herein, the term “promoter” or “transcription regulatory sequence” refers to a nucleic acid fragment that functions to control the transcription of one or more coding sequences, and is located upstream with respect to the direction of transcription of the transcription initiation site of the coding sequence, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA sequences, including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides known to one of skill in the art to act directly or indirectly to regulate the amount of transcription from the promoter. A “constitutive” promoter is a promoter that is active in most tissues under most physiological and developmental conditions. An “inducible” promoter is a promoter that is physiologically or developmentally regulated, e.g. by the application of a chemical inducer.

The term “selectable marker” is a term familiar to one of ordinary skill in the art and is used herein to describe any genetic entity which, when expressed, can be used to select for a cell or cells containing the selectable marker. The term “reporter” may be used interchangeably with marker, although it is mainly used to refer to visible markers, such as green fluorescent protein (GFP). Selectable markers may be dominant or recessive or bidirectional.

As used herein, the term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence. For instance, a transcription regulatory sequence is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein encoding regions, contiguous and in reading frame.

The terms “protein” or “polypeptide” are used interchangeably and refer to molecules consisting of a chain of amino acids, without reference to a specific mode of action, size, 3-dimensional structure or origin.

“Fungi” (singular fungus) are herein understood as heterotrophic eukaryotic microorganism that digest their food externally, absorbing nutrient molecules into their cells. Fungi are a separate kingdom of eukaryotic organisms and include yeasts, molds, and mushrooms. The terms fungi, fungus and fungal as used herein thus expressly includes yeasts as well as filamentous fungi.

The term “gene” means a DNA fragment comprising a region (transcribed region), which is transcribed into an RNA molecule (e.g. an mRNA) in a cell, operably linked to suitable regulatory regions (e.g. a promoter). A gene will usually comprise several operably linked fragments, such as a promoter, a 5’ leader sequence, a coding region and a 3’ nontranslated sequence (3’ end) comprising a polyadenylation site. “Expression of a gene” refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly a promoter, is transcribed into an RNA, which is biologically active, i.e. which is capable of being translated into a biologically active protein or peptide.

The term “homologous” when used to indicate the relation between a given (recombinant) nucleic acid or polypeptide molecule and a given host organism or host cell, is understood to mean that in nature the nucleic acid or polypeptide molecule is

produced by a host cell or organisms of the same species, preferably of the same variety or strain. If homologous to a host cell, a nucleic acid sequence encoding a polypeptide will typically (but not necessarily) be operably linked to another (heterologous) promoter sequence and, if applicable, another (heterologous) secretory
5 signal sequence and/or terminator sequence than in its natural environment. It is understood that the regulatory sequences, signal sequences, terminator sequences, etc. may also be homologous to the host cell. In this context, the use of only "homologous" sequence elements allows the construction of "self-cloned" genetically modified organisms (GMO's) (self-cloning is defined herein as in European Directive 98/81/EC
10 Annex II). When used to indicate the relatedness of two nucleic acid sequences the term "homologous" means that one single-stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt
15 concentration as discussed later.

The terms "heterologous" and "exogenous" when used with respect to a nucleic acid (DNA or RNA) or protein refers to a nucleic acid or protein that does not occur naturally as part of the organism, cell, genome or DNA or RNA sequence in which it is present, or that is found in a cell or location or locations in the genome or DNA or
20 RNA sequence that differ from that in which it is found in nature. Heterologous and exogenous nucleic acids or proteins are not endogenous to the cell into which it is introduced, but have been obtained from another cell or synthetically or recombinantly produced. Generally, though not necessarily, such nucleic acids encode proteins, i.e. exogenous proteins, that are not normally produced by the cell in which the DNA is
25 transcribed or expressed. Similarly exogenous RNA encodes for proteins not normally expressed in the cell in which the exogenous RNA is present. Heterologous/exogenous nucleic acids and proteins may also be referred to as foreign nucleic acids or proteins. Any nucleic acid or protein that one of skill in the art would recognize as foreign to the cell in which it is expressed is herein encompassed by the term heterologous or
30 exogenous nucleic acid or protein. The terms heterologous and exogenous also apply to non-natural combinations of nucleic acid or amino acid sequences, i.e. combinations where at least two of the combined sequences are foreign with respect to each other.

The “specific activity” of an enzyme is herein understood to mean the amount of activity of a particular enzyme per amount of total host cell protein, usually expressed in units of enzyme activity per mg total host cell protein. In the context of the present invention, the specific activity of a particular enzyme may be increased or decreased as compared to the specific activity of that enzyme in an (otherwise identical) wild type host cell.

“Anaerobic conditions” or an anaerobic fermentation process is herein defined as conditions or a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than 5, 2.5 or 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both electron donor and electron acceptors.

Detailed description of the invention

Biodiesel is having and will continue to have a substantial impact in the area of biofuels. It can be produced from a variety of sources including algae, rape and palm. A comprehensive list of potential feedstock sources is available from “Feedstock and Biodiesel Characteristics Report, “Renewable Energy Group, Inc., www.regfuel.com (2009)”. The vegetable oils from which the biodiesel is produced by interesterification inevitably result in large byproduct streams of glycerol. Furthermore, the plant material from which the oil is extracted will give rise to lignocellulosic byproduct. The latter by-product will contain cellulose and hemicelluloses in varying relative concentrations. These polymeric sugars can be hydrolyzed into glucose and pentoses such as xylose, respectively, by known processes that are still optimized. Hence, two by-product streams may result from the production of biodiesel. The present invention describes a method for the concomitant fermentation of both glycerol and the sugars into ethanol. This process can be applied for various feedstocks. As an example, the situation in the palm oil industry is presented in some detail below.

The palm oil industry produces huge amounts of by-products. The oil consists of only 10% of the total biomass produced in the plantation. The remainder consists of lignocellulosic materials such as oil palm fronds, trunks and empty fruit bunches (EFB) (See e.g. http://umpir.ump.edu.my/697/1/Kamarul_Azlan_Abd._Hamid.pdf). These lignocellulosic materials can be used for the production of value added products such as ethanol that may be used as biofuel. The lignocellulosic materials may be hydrolysed

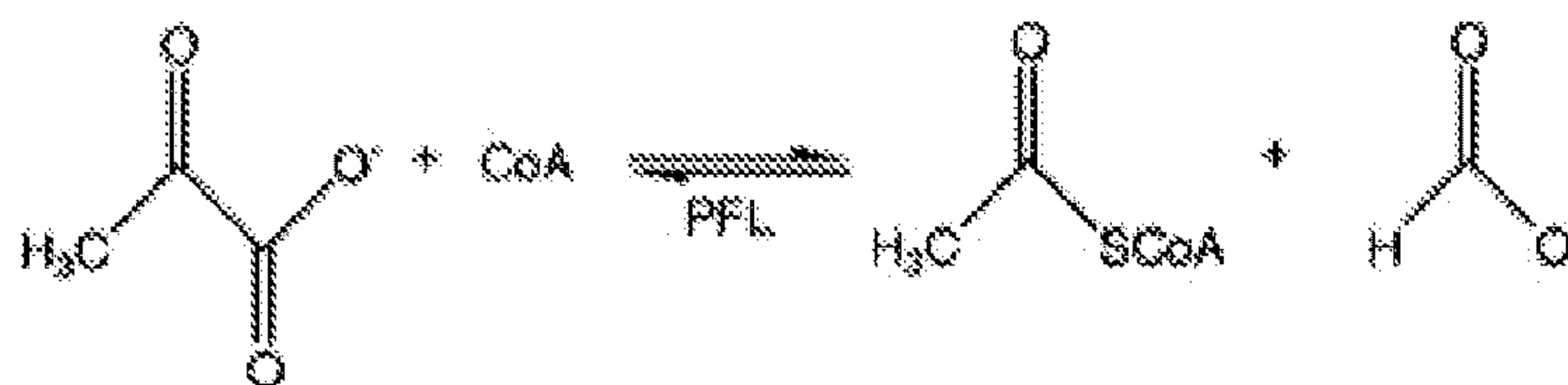
by known methods into streams containing hexoses and pentose such as glucose and xylose, which can then be fermented to ethanol by engineered yeast strains as e.g. described by Kuypers et al (2005, FEMS Yeast Res. 5:399-409). However, hexose- and pentose-containing streams obtained from hydrolysis of this type of lignocellulosic materials are relatively dilute, and will therefore yield relatively low ethanol concentrations of no more than about 5% (v/v).

At the same time, large amounts of concentrated glycerol are generated as a by-product in biodiesel production from transesterification reactions using vegetable oils such as palm oil and alcohols. The availability of crude glycerol is predicted to increase over the next years as a result of the growth in biodiesel production worldwide. Consequently large amounts of concentrated glycerol will be available at low cost near palm oil plantations. This offers the possibility to increase the carbon concentration of diluted streams obtained in the above-described hydrolysis of lignocellulosic materials by mixing in the available highly concentrated glycerol. Utilisation of such mixed streams would require engineered yeast strains which can ferment not only hexoses and pentose but also glycerol to ethanol. However, anaerobic consumption of large amount of glycerol would produce a redox imbalance in yeast. The present invention addresses this problem by engineering the yeast to produce formic acid, in addition to ethanol. The present invention therefore provides yeast strains engineered to produce ethanol and formic acid from carbon sources containing one or more of glycerol, hexose and pentose, as well as processes wherein these strains are used to produce ethanol and formic acid from these carbon sources.

In a first aspect the invention relates to a fungal host cell comprising an exogenous gene coding for a enzyme with the ability to convert pyruvate and coenzyme-A into formate and acetyl-CoA. An enzyme with the ability to convert pyruvate and coenzyme-A into formate and acetyl-CoA katalyses the reaction (EC 2.3.1.54):



11



Such an enzyme is herein understood as an enzyme having pyruvate formate lyase activity and is referred to as a pyruvate formate lyase (PFL) or formate C-acetyltransferase.

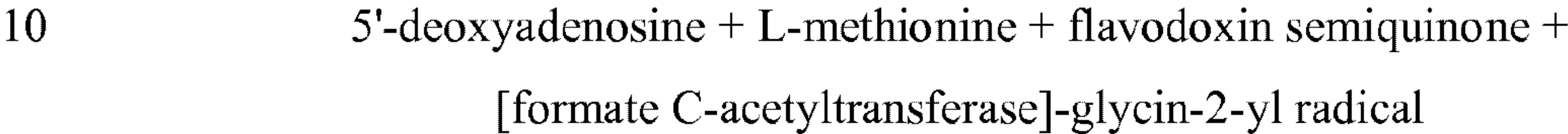
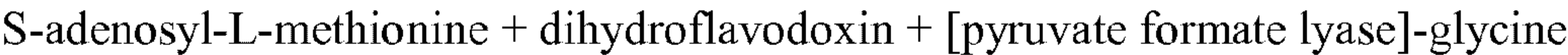
- 5 A suitable exogenous gene coding for an enzyme with pyruvate formate lyase activity is e.g. a prokaryotic pyruvate formate lyase, such as the pyruvate formate lyase from *E. coli*. The *E. coli* pyruvate formate lyase is a dimer of PflB (encoded by *pflB*), whose maturation requires the activating enzyme PflAE (encoded by *pflA*), radical S-adenosylmethionine, and a single electron donor, which in the case of *E. coli* is
- 10 flavodoxin (Buis and Broderick, 2005, Arch. Biochem. Biophys. 433:288–296; Sawers and Watson, 1998, Mol. Microbiol. 29:945–954). However, Waks and Silver (*supra*) have shown that for activation of the pyruvate formate lyase in yeast, only co-expression of an activating enzyme is required but expression of flavodoxin is not necessary.
- 15 The exogenous gene coding for an enzyme with pyruvate formate lyase activity preferably comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 1. Suitable examples of organisms comprising an enzyme with pyruvate formate lyase activity are provided in Table 1. Further examples of such organisms are
- 20 listed by Lehtiö and Goldman (2004, Prot. Engin. Design & Selection, 17:545-552). The amino acid sequences of these enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding enzyme with pyruvate formate lyase activity (see e.g. SEQ ID NO: 2). The exogenous gene coding for an enzyme with pyruvate formate lyase activity
- 25 may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 1. Preferably the amino acid sequence has no more than 420, 380, 300, 250, 200, 150, 100, 75, 50, 40, 30, 20, 10 or 5 amino acid substitutions, insertions and/or deletions as compared to SEQ ID NO: 1.

Table 1: Enzymes with pyruvate formate lyase activity related to *E.coli* pflB

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella boydii</i>	100%
<i>Escherichia albertii</i> TW07627	99%
<i>Salmonella enterica</i>	97%
<i>Citrobacter rodentium</i> ICC168	97%
<i>Klebsiella pneumoniae</i> NTUH-K2044	96%
<i>Yersinia aldovae</i> ATCC 35236	91%
<i>Proteus mirabilis</i> HI4320	87%
<i>Haemophilus influenzae</i> Rd KW20	86%
<i>Actinobacillus succinogenes</i> 130Z	83%
<i>Piromyces</i> sp. E2	57%

The host cell of the invention further preferably comprises an exogenous gene coding for the PflAE activating enzyme for activation of the pyruvate formate lyase.

5 The pyruvate formate lyase activating enzyme is herein understood as an enzyme that katalyses the reaction:



The exogenous gene coding for the pyruvate formate lyase activating enzyme preferably comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 3. Suitable examples of organisms comprising an enzyme with pyruvate formate lyase activity are provided in Table 2. The amino acid sequences of these enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding enzyme with pyruvate formate lyase activity (see e.g. SEQ ID NO: 4). The exogenous gene

15

20

coding for the pyruvate formate lyase activating enzyme may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 3. Preferably the amino acid sequence has no more than 135, 125, 100,
5 75, 50, 25, 20, 15, 10, 8, 5 or 2 amino acid substitutions, insertions and/or deletions as compared to SEQ ID NO: 3.

Table 2: Pyruvate formate lyase activating enzymes related to *E.coli* pflA

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella boydii</i>	100%
<i>Escherichia albertii</i> TW07627	99%
<i>Salmonella enterica</i>	98%
<i>Citrobacter rodentium</i> ICC168	98%
<i>Klebsiella pneumoniae</i> NTUH-K2044	97%
<i>Yersinia rohdei</i> ATCC 43380	89%
<i>Proteus penneri</i> ATCC 35198	85%
<i>Haemophilus parasuis</i> 29755	70%

10 In a preferred host cell of the invention, the exogenous genes coding for the enzyme with pyruvate formate lyase activity and the pyruvate formate lyase activating enzyme are from the same donor organism, i.e. be homologous to each other. However, the exogenous genes coding for the enzyme with pyruvate formate lyase activity and the pyruvate formate lyase activating enzyme may also be from different donor
15 organisms, i.e. be heterologous to each other.

In one aspect the invention relates to methods for preparing or constructing the yeast cells of the invention. For this purpose standard genetic and molecular biology techniques are used that are generally known in the art and have e.g. been described by Sambrook and Russell (2001, "Molecular cloning: a laboratory manual" (3rd edition),
20 Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press) and Ausubel et al. (1987, eds., "Current protocols in molecular biology", Green Publishing and Wiley

Interscience, New York). Furthermore, the construction of mutated host yeast strains is carried out by genetic crosses, sporulation of the resulting diploids, tetrad dissection of the haploid spores containing the desired auxotrophic markers, and colony purification of such haploid host yeasts in the appropriate selection medium. All of these methods
5 are standard yeast genetic methods known to those in the art. See, for example, Sherman et al., *Methods Yeast Genetics*, Cold Spring Harbor Laboratory, NY (1978) and Guthrie et al. (Eds.) *Guide To Yeast Genetics and Molecular Biology* Vol. 194, Academic Press, San Diego (1991).

The exogenous genes coding for the enzyme having pyruvate formate lyase
10 activity and the pyruvate formate lyase activating enzyme, preferably are expression constructs comprising the nucleotide sequence coding for the enzymes operably linked to suitable expression regulatory regions/sequences to ensure expression of the enzymes upon transformation of the expression constructs into the host cell of the invention. Thus, the gene or expression construct will at least comprise a promoter that
15 is functional in the host cell operably linked to the coding sequence. The gene or construct may further comprise a 5' leader sequence upstream of the coding region and a 3'-nontranslated sequence (3'end) comprising a polyadenylation site and a transcription termination site downstream of the coding sequence. It is understood that the nucleotide sequences coding for the enzyme having pyruvate formate lyase activity
20 and the pyruvate formate lyase activating enzyme may be present together on a single expression construct, or each enzyme may be present on a separate expression construct.

Suitable promoters for expression of the nucleotide sequences coding for the enzyme having pyruvate formate lyase activity and the pyruvate formate lyase
25 activating enzyme (as well as other enzymes of the invention; see below) include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Promoters having these characteristics are widely available and known to the skilled person. Suitable examples of such promoters include e.g.
30 promoters from glycolytic genes such as the phosphofructokinase (*PPK*), triose phosphate isomerase (*TPI*), glyceraldehyde-3-phosphate dehydrogenase (*GPD*, *TDH3* or *GAPDH*), pyruvate kinase (*PYK*), phosphoglycerate kinase (*PGK*), glucose-6-phosphate isomerase promoter (*PGII*) promoters from yeasts. More details about such

promoters from yeast may be found in (WO 93/03159). Other useful promoters are ribosomal protein encoding gene promoters (*TEFI*), the lactase gene promoter (*LAC4*), alcohol dehydrogenase promoters (*ADHI*, *ADH4*, and the like), the enolase promoter (*ENO*) and the hexose(glucose) transporter promoter (*HXT7*). Alternatively, the
5 nucleotide sequences encoding the enzyme having pyruvate formate lyase activity and the PflAE activating enzyme are expressed under anaerobic conditions by using an anoxic promoter such as e.g. the *S. cerevisiae ANB1* promoter (SEQ ID NO: 24). Other promoters, both constitutive and inducible, and enhancers or upstream activating sequences will be known to those of skill in the art. Preferably the promoter that is
10 operably linked to nucleotide sequence as defined above is homologous to the host cell. Suitable terminator sequences are e.g. obtainable from the cytochrome c1 (*CYC1*) gene or an alcohol dehydrogenase gene (e.g. *ADHI*).

To increase the likelihood that the enzyme having pyruvate formate lyase activity is expressed at sufficient levels and in active form in the transformed host cells of the
15 invention, the nucleotide sequence encoding these enzymes, as well as the pyruvate formate lyase activating enzyme and other enzymes of the invention (see below), are preferably adapted to optimise their codon usage to that of the host cell in question. The adaptiveness of a nucleotide sequence encoding an enzyme to the codon usage of a host cell may be expressed as codon adaptation index (CAI). The codon adaptation index is
20 herein defined as a measurement of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes in a particular host cell or organism. The relative adaptiveness (w) of each codon is the ratio of the usage of each codon, to that of the most abundant codon for the same amino acid. The CAI index is defined as the geometric mean of these relative adaptiveness values. Non-synonymous
25 codons and termination codons (dependent on genetic code) are excluded. CAI values range from 0 to 1, with higher values indicating a higher proportion of the most abundant codons (see Sharp and Li, 1987, Nucleic Acids Research 15: 1281-1295; also see: Jansen et al., 2003, Nucleic Acids Res. 31(8):2242-51). An adapted nucleotide sequence preferably has a CAI of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 or 0.9. Most
30 preferred are the sequences which have been codon optimised for expression in the fungal host cell in question such as e.g. *S. cerevisiae* cells.

The fungal host cell to be transformed with a nucleic acid construct comprising a nucleotide sequence encoding an enzyme with pyruvate formate lyase activity

preferably is a yeast host cell. Preferably the host cell is a cultured cell. The host cell of the invention, preferably is a host capable of active or passive pentose (xylose and preferably also arabinose) transport into the cell. The host cell preferably contains active glycolysis. The host cell may further preferably contains an endogenous pentose phosphate pathway and may contain endogenous xylulose kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. The host further preferably contains enzymes for conversion of a pentose (preferably through pyruvate) to a desired fermentation product such as ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products. A particularly preferred host cell is a yeast cell that is naturally capable of alcoholic fermentation, preferably, anaerobic alcoholic fermentation. The yeast host cell further preferably has a high tolerance to ethanol, a high tolerance to low pH (i.e. capable of growth at a pH lower than 5, 4, or 3) and towards organic acids like lactic acid, acetic acid or formic acid and sugar degradation products such as furfural and hydroxy-methylfurfural, and a high tolerance to elevated temperatures. Any of these characteristics or activities of the host cell may be naturally present in the host cell or may be introduced or modified by genetic modification, preferably by self cloning or by the methods of the invention described below. A suitable cell is a cultured cell, a cell that may be cultured in fermentation process e.g. in submerged or solid state fermentation. Particularly suitable host cells are eukaryotic microorganism like e.g. fungi, however, most suitable for use in the present inventions are yeasts.

Yeasts are herein defined as eukaryotic microorganisms and include all species of the subdivision Eumycotina (Yeasts: characteristics and identification, J.A. Barnett, R.W. Payne, D. Yarrow, 2000, 3rd ed., Cambridge University Press, Cambridge UK; and, The yeasts, a taxonomic study, C.P. Kurtzman and J.W. Fell (eds) 1998, 4th ed., Elsevier Science Publ. B.V., Amsterdam, The Netherlands) that predominantly grow in unicellular form. Yeasts may either grow by budding of a unicellular thallus or may grow by fission of the organism. Preferred yeasts cells for use in the present invention belong to the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*. Preferably the yeast is capable of anaerobic fermentation, more preferably anaerobic alcoholic fermentation. Over the years suggestions have been made for the introduction

of various organisms for the production of bio-ethanol from crop sugars. In practice, however, all major bio-ethanol production processes have continued to use the yeasts of the genus *Saccharomyces* as ethanol producer. This is due to the many attractive features of *Saccharomyces* species for industrial processes, i.e., a high acid-, ethanol-
5 and osmo-tolerance, capability of anaerobic growth, and of course its high alcoholic fermentative capacity. Preferred yeast species as host cells include *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.

In a further embodiment, the host cell of the invention further comprises a genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in
10 the cell. NAD^+ -dependent formate dehydrogenases (FDH; EC 1.2.1.2) catalyse the oxidation of formate to bicarbonate, donating the electrons to NAD^+ . In the cells of the invention, the specific formate dehydrogenase activity is preferably reduced by at least a factor 0.8, 0.5, 0.3, 0.1, 0.05 or 0.01 as compared to a strain which is genetically identical except for the genetic modification causing the reduction in specific activity,
15 preferably under anaerobic conditions. Formate dehydrogenase activity may be determined as described by Overkamp et al. (2002, Yeast 19:509-520).

Preferably, formate dehydrogenase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an formate dehydrogenase. Preferably, the genetic modifications reduce or inactivate
20 the expression of each endogenous copy of the gene encoding a specific formate dehydrogenase in the cell's genome. A given cell may comprise multiple copies of the gene encoding a specific formate dehydrogenase with one and the same amino acid sequence as a result of di-, poly- or aneu-ploidy. In such instances preferably the expression of each copy of the specific gene that encodes the formate dehydrogenase is
25 reduced or inactivated. Alternatively, a cell may contain several different (iso)enzymes with formate dehydrogenase activity that differ in amino acid sequence and that are each encoded by a different gene. In such instances, in some embodiments of the invention it may be preferred that only certain types of the isoenzymes are reduced or inactivated while other types remain unaffected. Preferably, however, expression of all
30 copies of genes encoding (iso)enzymes with formate dehydrogenase activity is reduced or inactivated.

Preferably, a gene encoding formate dehydrogenase activity is inactivated by deletion of at least part of the gene or by disruption of the gene, whereby in this context

the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of formate dehydrogenase activity in the host cell.

A preferred gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is the *S. cerevisiae FDH1* as
5 described by van den Berg and Steensma (1997, Yeast 13:551-559), encoding the amino acid sequence of SEQ ID NO: 5 and orthologues thereof in other species. Therefore a gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention preferably is a gene encoding a formate
10 dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 5 or a gene encoding a formate dehydrogenase having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 5.

However, in some strains of *S. cerevisiae* a second gene encoding a formate
15 dehydrogenase is active, i.e. the *FDH2*, see e.g. Overkamp et al. (2002, *supra*). Another preferred gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention therefore is an *S. cerevisiae FDH2* as described by Overkamp et al. (2002, *supra*), encoding the amino acid sequence of SEQ ID NO: 6 and orthologues thereof in other species. Therefore a gene encoding a formate
20 dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention preferably is a gene encoding a formate dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 6 or a gene encoding a formate dehydrogenase having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the
25 amino acid sequence of SEQ ID NO: 6.

In a further embodiment, the activity of all the genes in the host cell encoding a formate dehydrogenase is reduced or inactivated. In such cells preferably all copies of endogenous genes encoding a formate dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ
30 ID NO: 5 or 6 (or having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequences of SEQ ID NO: 5 or 6) are inactivated or at least reduced in expression.

In a preferred embodiment, the host cell of the invention further comprises an exogenous gene coding for an enzyme with the ability to reduce acetylCoA into acetaldehyde, which gene confers to the cell the ability to convert acetylCoA (and/or acetic acid) into ethanol. An enzyme with the ability to reduce acetylCoA into
5 acetaldehyde is herein understood as an enzyme which catalyze the reaction (ACDH; EC 1.2.1.10):



10 Thus, the enzyme catalyzes the conversion of acetylCoA into acetaldehyde (and *vice versa*) and is also referred to as an (acetylating NAD-dependent) acetaldehyde dehydrogenase or an acetyl-CoA reductase. The enzyme may be a bifunctional enzyme which further catalyzes the conversion of acetaldehyde into ethanol (and *vice versa*; see below). For convenience we shall refer herein to an enzyme having at least the ability
15 to reduce acetylCoA into either acetaldehyde or ethanol as an “acetaldehyde dehydrogenase”. It is further understood herein that the host cell has endogenous alcohol dehydrogenase activities which allow the cell, being provided with acetaldehyde dehydrogenase activity, to complete the conversion of acetyl-CoA into ethanol. It is further also preferred that the host cell has endogenous acetyl-CoA
20 synthetase which allow the cell, being provided with acetaldehyde dehydrogenase activity, to complete the conversion of acetic acid (via acetyl-CoA) into ethanol.

The exogenous gene may encode for a monofunctional enzyme having only acetaldehyde dehydrogenase activity (i.e. an enzyme only having the ability to reduce acetylCoA into acetaldehyde) such as e.g. the acetaldehyde dehydrogenase encoded by
25 the *E.coli* mhpF gene. A suitable exogenous gene coding for an enzyme with acetaldehyde dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 7. Suitable examples of prokaryotes comprising monofunctional enzymes with acetaldehyde dehydrogenase activity are
30 provided in Table 3. The amino acid sequences of these monofunctional enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding monofunctional enzyme (see e.g. SEQ ID NO: 8). The exogenous gene coding for the monofunctional enzyme

having only acetaldehyde dehydrogenase activity may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 7.

5 Table 3: Enzymes with acetaldehyde dehydrogenase activity related to *E.coli* mhpF

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella sonnei</i>	100%
<i>Escherichia coli</i> IAI39	99%
<i>Citrobacter youngae</i> ATCC 29220	93%
<i>Citrobacter</i> sp. 30_2	92%
<i>Klebsiella pneumoniae</i> 342)	87%
<i>Klebsiella variicola</i>	87%
<i>Pseudomonas putida</i>	81%
<i>Ralstonia eutropha</i> JMP134	82%
<i>Burkholderia</i> sp. H160	81%
<i>Azotobacter vinelandii</i> DJ	79%
<i>Ralstonia metallidurans</i> CH34	70%
<i>Xanthobacter autotrophicus</i> Py2	67%
<i>Burkholderia cenocepacia</i> J2315	68%
<i>Frankia</i> sp. EAN1pec	67%
<i>Polaromonas</i> sp. JS666	68%
<i>Burkholderia phytofirmans</i> PsJN	70%
<i>Rhodococcus opacus</i> B4	64%

Preferably, the host cell comprises an exogenous gene coding for a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity, which gene confers to the cell the ability to convert acetylCoA into ethanol. The advantage of using a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activities as opposed to separate enzymes for each of the acetaldehyde dehydrogenase and alcohol dehydrogenase activities, is that it allows for direct channelling of the intermediate between enzymes that catalyze consecutive reactions in

a pathway offers the possibility of an efficient, exclusive, and protected means of metabolite delivery. Substrate channelling thus decreases transit time of intermediates, prevents loss of intermediates by diffusion, protects labile intermediates from solvent, and forestalls entrance of intermediates into competing metabolic pathways. The bifunctional enzyme therefore allows for a more efficient conversion of acetylCoA into ethanol as compared to the separate acetaldehyde dehydrogenase and alcohol dehydrogenase enzymes. A further advantage of using the bifunctional enzyme is that it may also be used in host cells having little or no alcohol dehydrogenase activity under the condition used, such as e.g. anaerobic conditions and/or conditions of catabolite repression.

Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are known in the art prokaryotes and protozoans, including e.g. the bifunctional enzymes encoded by the *Escherichia coli* adhE and *Entamoeba histolytica* ADH2 genes (see e.g. Bruchaus and Tannich, 1994, J. Biochem. 303: 743-748; Burdette and Zeikus, 1994, J. Biochem. 302: 163-170; Koo et al., 2005, Biotechnol. Lett. 27: 505-510; Yong et al., 1996, Proc Natl Acad Sci USA, 93: 6464-6469). Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are larger proteins consisting of around 900 amino acids and they are bifunctional in that they exhibit both acetaldehyde dehydrogenase (ACDH; EC 1.2.1.10) and alcohol dehydrogenase activity (ADH; EC 1.1.1.1). The *E. coli* adhE and *Entamoeba histolytica* ADH2 show 45% amino acid identity. Therefore, in one embodiment of the invention, a suitable exogenous gene coding for a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO: 9 and 11. Suitable examples of prokaryotes comprising bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are provided in Tables 4 and 5. The amino acid sequences of these bifunctional enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding bifunctional enzyme (see e.g. SEQ ID NO: 10 or 12). The exogenous gene coding for the a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity may also comprises a nucleotide sequence coding for an amino

acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO: 9 and 11.

5 Table 4: Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity related to *E.coli* adhE

Organism	Amino acid identity (%)
<i>Escherichia coli</i> O157:H7 str. Sakai	100%
<i>Shigella sonnei</i>	100%
<i>Shigella dysenteriae</i> 1012	99%
<i>Klebsiella pneumoniae</i> 342	97%
<i>Enterobacter</i> sp. 638	94%
<i>Yersinia pestis</i> biovar Microtus str. 91001	90%
<i>Serratia proteamaculans</i> 568	90%
<i>Pectobacterium carotovorum</i> WPP14	90%
<i>Sodalis glossinidius</i> str. 'morsitans'	87%
<i>Erwinia tasmaniensis</i> Et1/99	86%
<i>Aeromonas hydrophila</i> ATCC 7966	81%
<i>Vibrio vulnificus</i> YJ016]	76%

Table 5: Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity related to *Entamoeba histolytica* ADH2

Organism	Amino acid identity (%)
<i>Entamoeba histolytica</i> HM-1:IMSS	99%
<i>Entamoeba dispar</i> SAW760	98%
<i>Mollicutes bacterium</i> D7	65%
<i>Fusobacterium mortiferum</i> ATCC 9817	64%
<i>Actinobacillus succinogenes</i> 130Z	63%
<i>Pasteurella multocida</i> Pm70	62%
<i>Mannheimia succiniciproducens</i> MBEL55E	61%
<i>Streptococcus</i> sp. 2_1_36FAA]	61%

For expression of the nucleotide sequence encoding the bifunctional enzyme having acetaldehyde dehydrogenase and alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity, the nucleotide sequence (to be
5 expressed) is placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure expression of the enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for expression of the nucleotide sequence coding for the enzyme having the bifunctional enzyme having acetaldehyde dehydrogenase and
10 alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above.

Preferably, the nucleotide sequence encoding the bifunctional enzyme having
15 acetaldehyde dehydrogenase and alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity is adapted to optimise its codon usage to that of the host cell in question (as described above).

The enzyme having acetaldehyde dehydrogenase and optionally alcohol dehydrogenase activities preferably is expressed in active form in the transformed host
20 cell. Thus, expression of the nucleotide sequence in the host cell produces an acetaldehyde dehydrogenase with a specific activity of at least 0.005, 0.010, 0.020, 0.050 or 0.10 $\mu\text{mol min}^{-1} (\text{mg protein})^{-1}$, determined as acetyl-CoA dependent rate of NADH reduction in cell extracts of the transformed host cell at 30 °C as described in the Examples herein.

25 In a further embodiment, the host cell of the invention further comprises a genetic modification that increases at least one of: i) the specific activity of glycerol dehydrogenase; ii) the specific activity of dihydroxyacetone kinase; and, iii) transport of glycerol into the cell.

Preferably, the genetic modification that increases the specific activity of at least
30 one of glycerol dehydrogenase and dihydroxyacetone kinase is overexpression of a nucleotide sequence encoding at least one of a glycerol dehydrogenase and dihydroxyacetone kinase. However, alternatively, the specific activity of the glycerol dehydrogenase and/or dihydroxyacetone kinase may be increased by expressing an

enzyme having increased activity as compared to the endogenous wild type enzyme of the host cell, in addition to, or as a replacement for the wild type enzyme.

A glycerol dehydrogenase is herein understood as an enzyme that catalyzes the chemical reaction (EC 1.1.1.6):

5



Other names in common use include glycerin dehydrogenase, NAD^+ -linked glycerol dehydrogenase and glycerol: NAD^+ 2-oxidoreductase. Preferably the genetic
10 modification causes overexpression of a glycerol dehydrogenase, e.g. by overexpression of a nucleotide sequence encoding a glycerol dehydrogenase. The nucleotide sequence encoding the glycerol dehydrogenase may be endogenous to the cell or may be a glycerol dehydrogenase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of glycerol dehydrogenase in the cells
15 of the invention are e.g. the glycerol dehydrogenase gene from *S. cerevisiae* (*GCY1*) as e.g. described by Oechsner et al. (1988, FEBS Lett. 238: 123-128) or Voss et al. (1997, Yeast 13: 655-672). Preferably, the nucleotide sequence encoding the glycerol dehydrogenase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence
20 identity with SEQ ID NO: 13 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 13. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol dehydrogenase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the glycerol dehydrogenase of SEQ ID NO:
25 13.

The glycerol dehydrogenase encoded by the yeast *GCY1* gene however appears to be specific for the cofactor NADP^+ (EC 1.1.1.72) as opposed to NAD^+ (EC 1.1.1.6). The yeasts such *S. cerevisiae* appear to lack NAD^+ -dependent glycerol dehydrogenase activity (EC 1.1.1.6) (see e.g. KEGG pathway 00561). More preferred nucleotide
30 sequences for overexpression of a heterologous glycerol dehydrogenase in the cells of the invention are therefore e.g. sequences encoding bacterial glycerol dehydrogenases which use NAD^+ as cofactor (EC 1.1.1.6), such as e.g. the *gldA* gene from *E. coli* described by Truniger and Boos (1994, J Bacteriol. 176(6):1796-1800), the expression

of which in yeast has already been reported (Lee and Dasilva, 2006, Metab Eng. 8(1):58-65). Preferably, the nucleotide sequence encoding a heterologous glycerol dehydrogenase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence
5 identity with SEQ ID NO: 49 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 49. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the heterologous glycerol dehydrogenase is overexpressed, such as e.g. a codon-optimised nucleotide sequence encoding the amino acid sequence of the
10 glycerol dehydrogenase of SEQ ID NO: 49. Such a codon-optimised nucleotide sequence is e.g. provided in SEQ ID NO: 50 (positions 10 - 1113; CAI = 0.976).

For overexpression of the nucleotide sequence encoding the glycerol dehydrogenase, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory
15 regions/sequences to ensure overexpression of the glycerol dehydrogenase enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the enzyme having glycerol dehydrogenase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic
20 conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above. In the cells of the invention, a glycerol dehydrogenase to be overexpressed is preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the glycerol
25 dehydrogenase is overexpressed under anaerobic conditions by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the
30 steady state level of the transcript coding for the enzyme in the cell. Overexpression of the nucleotide sequence in the host cell produces a specific glycerol dehydrogenase activity of at least 0.2, 0.5, 1.0, 2.0, or 5.0 U min⁻¹ (mg protein)⁻¹, determined in cell extracts of the transformed host cells at 30 °C as described in the Examples herein.

A dihydroxyacetone kinase is herein understood as an enzyme that catalyzes the chemical reaction ((EC 2.7.1.29):



5

Other names in common use include glycerone kinase, ATP:glycerone phosphotransferase and (phosphorylating) acetol kinase. It is understood that glycerone and dihydroxyacetone are the same molecule. Preferably the genetic modification causes overexpression of a dihydroxyacetone kinase, e.g. by overexpression of a nucleotide sequence encoding a dihydroxyacetone kinase. The nucleotide sequence encoding the dihydroxyacetone kinase may be endogenous to the cell or may be a dihydroxyacetone kinase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of dihydroxyacetone kinase in the cells of the invention are e.g. the dihydroxyacetone kinase genes from *S. cerevisiae* (*DAK1*) and (*DAK2*) as e.g. described by Molin et al. (2003, J. Biol. Chem. 278:1415-1423). Preferably, the nucleotide sequence encoding the dihydroxyacetone kinase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 14 and 15 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 14 and 15. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the dihydroxyacetone kinase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the dihydroxyacetone kinase of SEQ ID NO: 14 or a codon optimised nucleotide sequence encoding the dihydroxyacetone kinase of SEQ ID NO: 15. A preferred nucleotide sequence for overexpression of a dihydroxyacetone kinase is a nucleotide sequence encoding a dihydroxyacetone kinase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO's: 14 (*S. cerevisiae* (*DAK1*), having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 14.

Nucleotide sequences that may be used for overexpression of a heterologous dihydroxyacetone kinase in the cells of the invention are e.g. sequences encoding bacterial dihydroxyacetone kinases such as the *dhaK* gene from *Citrobacter freundii* e.g. described by Daniel et al. (1995, J. Bacteriol. 177:4392-4401). Preferably, the

30

nucleotide sequence encoding a heterologous dihydroxyacetone kinase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 52 or a nucleotide sequence coding for an amino acid sequence having one or several
5 substitutions, insertions and/or deletions as compared to SEQ ID NO: 52. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the heterologous dihydroxyacetone kinase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the amino acid sequence of the dihydroxyacetone kinase of SEQ ID NO: 52. Such a codon-optimised nucleotide sequence is e.g. provided in
10 SEQ ID NO: 53 (positions 10 - 1668).

For overexpression of the nucleotide sequence encoding the dihydroxyacetone kinase, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the dihydroxyacetone kinase enzyme
15 upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the enzyme having dihydroxyacetone kinase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction.
20 Examples of such promoters are given above. In the cells of the invention, a dihydroxyacetone kinase to be overexpressed is preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the dihydroxyacetone kinase is overexpressed under anaerobic conditions by at least a
25 factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme in the cell.
30 Overexpression of the nucleotide sequence in the host cell produces a specific dihydroxyacetone kinase activity of at least 0.002, 0.005, 0.01, 0.02 or 0.05 U min⁻¹ (mg protein)⁻¹, determined in cell extracts of the transformed host cells at 30 °C as described in the Examples herein.

Preferably, the genetic modification that increases transport of glycerol into the cell preferably is a genetic modification that causes overexpression of a nucleotide sequence encoding at least one of a glycerol uptake protein and a glycerol channel.

A glycerol uptake protein is herein understood as a multimembrane-spanning protein that belongs to the included in the membrane bound O-acyltransferases (MBOAT) superfamily including e.g. the *S. cerevisiae* glycerol uptake proteins encoded by the *GUP1* and *GUP2* genes. Preferably the genetic modification causes overexpression of a glycerol uptake protein, e.g. by overexpression of a nucleotide sequence encoding a glycerol uptake protein. The nucleotide sequence encoding the glycerol uptake protein may be endogenous to the cell or may be a glycerol uptake protein that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of glycerol uptake protein in the cells of the invention are e.g. the glycerol uptake protein genes from *S. cerevisiae* (*GUP1*) and (*GUP2*) and orthologues thereof as e.g. described by Neves et al. (2004, FEMS Yeast Res. 5:51-62). Preferably, the nucleotide sequence encoding the glycerol uptake protein comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 16 (Gup1p) and 17 (Gup2p) or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 16 and 17. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol uptake protein is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the glycerol uptake protein SEQ ID NO: 16 or a codon optimised nucleotide sequence encoding the glycerol uptake protein of SEQ ID NO: 17. Although the exact nature of the influence of GUP1 on glycerol transport is not yet clear, Yu et al. (2010, *supra*) have shown that overexpression of GUP1 in *S. cerevisiae* improves ethanol production on glycerol grown cells. A preferred nucleotide sequence for overexpression of a glycerol uptake protein is therefore a nucleotide sequence encoding a glycerol uptake protein that is capable of rescuing salt stress-associated phenotype of a *S. cerevisiae gup1Δ* mutant by complementation as described by Neves et al. (2004, *supra*). Such complementing orthologues of *S. cerevisiae GUP1* include nucleotide sequences encoding amino acid sequences having at least 60, 68, 72, 75, 80, 85, 90, 95, 98, 99% identity with the amino acid sequence of SEQ ID NO: 16 and may be obtained from yeast species

belonging to the genera of *Saccharomyces*, *Zygosaccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*.

A glycerol channel is herein understood as a member of the MIP family of channel proteins reviewed by Reizer et al. (1993, CRC Crit. Rev. Biochem. Mol. Biol., 28: 235-257), which channel proteins comprise a 250 -280 amino acid transmembrane domain consisting of six membrane-spanning domains and have at least 30, 35, 40, 45, 50, 60, 70, 80, 90, 95, 98 or 99 % amino acid identity, or at least 55, 60, 65, 70, 80, 90, 95, 98 or 99% amino acid similarity with the amino acid sequence between amino acids 250 and 530 of SEQ ID NO: 18, the *S. cerevisiae* FPS1 aquaglyceroporin. Alternatively the channel protein comprises a 250 -280 amino acid transmembrane domain consisting of six membrane-spanning domains, and having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence between amino acids 250 and 530 of SEQ ID NO: 18.

Nucleotide sequences that may be used for overexpression of a glycerol channel in the cells of the invention include nucleotide sequences encoding the yeast aquaglyceroporin FPS1 gene from e.g. *S. cerevisiae* (Van Aelst et al., 1991, EMBO J. 10:2095-2104) and orthologues thereof from other yeasts including *Kluyveromyces lactis*, *Kluyveromyces marxianus* and *Zygosaccharomyces rouxii* as e.g. described by Neves et al. (2004, *supra*). However, the use of bacterial or plant glycerol channels is not excluded as e.g. Luyten et al. (1995, EMBO J. 14:1360-1371) have shown that the *E.coli* glycerol facilitator, having only 30% sequence identity with the amino acid sequence between amino acids 250 and 530 of the *S. cerevisiae* FPS1 aquaglyceroporin, can complement glycerol uptake in a *S. cerevisiae* *fps1Δ* mutant. The nucleotide sequence encoding the glycerol channel may be endogenous to the cell or may be a glycerol channel that is heterologous to the cell. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol channel is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the aquaglyceroporin of SEQ ID NO: 18.

For overexpression of the nucleotide sequence encoding the glycerol uptake protein and/or the glycerol channel protein, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the glycerol uptake protein and/or the glycerol channel protein upon transformation of the

expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the glycerol uptake protein and/or the glycerol channel protein include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above. In the cells of the invention, a glycerol uptake protein and/or a glycerol channel protein to be overexpressed are preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the glycerol uptake protein and/or the glycerol channel protein are overexpressed under anaerobic conditions by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme in the cell.

In a further embodiment, the host cell of the invention further comprises a genetic modification that increases the specific acetyl-CoA synthetase activity in the cell, preferably under anaerobic conditions as this activity is rate-limiting under these conditions. Acetyl-CoA synthetase or acetate-CoA ligase (EC 6.2.1.1) is herein understood as an enzyme that catalyzes the formation of a new chemical bond between acetate and coenzyme A (CoA). Preferably the genetic modification causes overexpression of a acetyl-CoA synthetase, e.g. by overexpression of a nucleotide sequence encoding a acetyl-CoA synthetase. The nucleotide sequence encoding the acetyl-CoA synthetase may be endogenous to the cell or may be a acetyl-CoA synthetase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of acetyl-CoA synthetase in the cells of the invention are e.g. the acetyl-CoA synthetase genes from *S. cerevisiae* (*ACS1* and *ACS2*) as e.g. described by de Jong-Gubbels et al. (1998, FEMS Microbiol Lett. 165: 15-20). Preferably, the nucleotide sequence encoding the acetyl-CoA synthetase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 19 and 20, or a nucleotide

sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 19 and 20.

In one embodiment, the nucleotide sequence that is overexpressed encodes an acetyl-CoA synthetase with a high affinity for acetate. Use of an acetyl-CoA synthetase
5 with a high affinity for acetate is preferred for conditions under which there is a relatively low concentration of acetic acid in the culture medium, e.g. no more than 2 g acetic acid/L culture medium. An acetyl-CoA synthetase with a high affinity for acetate is herein defined as an acetyl-CoA synthetase with a higher affinity for acetate than the acetyl-CoA synthetase encoded by the *S. cerevisiae ACS2* (SEQ ID NO: 20).
10 Preferably, an acetyl-CoA synthetase with a high affinity for acetate has a K_m for acetate of no more than 10, 5, 2, 1, 0.5, 0.2 or 0.1 mM, such e.g. the acetyl-CoA synthetase encoded by the *S. cerevisiae ACS1* gene. More preferably a codon-optimised (see above) nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 19 is overexpressed.

15 In another embodiment, the nucleotide sequence that is overexpressed encodes an acetyl-CoA synthetase with a high maximum rate (v_{max}). Use of an acetyl-CoA synthetase with a high maximum rate is preferred for condition under which there is a relatively high concentration of acetic acid in the culture medium, e.g. at least 2, 3, 4 or 5 g acetic acid/L culture medium. An acetyl-CoA synthetase with a high maximum rate
20 is herein defined as an acetyl-CoA synthetase with a higher maximum rate than the acetyl-CoA synthetase encoded by the *S. cerevisiae ACS1*. Preferably, the acetyl-CoA synthetase with a high maximum rate is the acetyl-CoA synthetase encoded by the *S. cerevisiae ACS2* gene. More preferably a codon-optimised (see above) nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 20 is overexpressed.

25 For overexpression of the nucleotide sequence encoding the acetyl-CoA synthetase (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the acetyl-CoA synthetase enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters
30 for (over)expression of the nucleotide sequence coding for the enzyme having acetyl-CoA synthetase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given

above. In the cells of the invention, an acetyl-CoA synthetase to be overexpressed is overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the acetyl-CoA synthetase is overexpressed under
5 anaerobic conditions by at least a factor 2, 5, 10, 20, 50, or 100 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the
10 enzyme.

In a further embodiment, the host cell of the invention further comprises a genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell. Glycerol 3-phosphate dehydrogenase or glycerolphosphate dehydrogenase (EC 1.1.1.8) katalyses the reduction of
15 dihydroxyacetone phosphate to sn-glycerol 3-phosphate while oxidising NADH to NAD^+ . In the cells of the invention, the specific glycerolphosphate dehydrogenase activity is preferably reduced by at least a factor 0.8, 0.5, 0.3, 0.1, 0.05 or 0.01 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression, preferably under anaerobic conditions.

20 Preferably, glycerolphosphate dehydrogenase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an glycerolphosphate dehydrogenase. Preferably, the genetic modifications reduce or inactivate the expression of each endogenous copy of the gene encoding a specific glycerolphosphate dehydrogenase in the cell's genome. A given cell may
25 comprise multiple copies of the gene encoding a specific glycerolphosphate dehydrogenase with one and the same amino acid sequence as a result of di-, poly- or aneu-ploidy. In such instances preferably the expression of each copy of the specific gene that encodes the glycerolphosphate dehydrogenase is reduced or inactivated. Alternatively, a cell may contain several different (iso)enzymes with glycerolphosphate
30 dehydrogenase activity that differ in amino acid sequence and that are each encoded by a different gene. In such instances, in some embodiments of the invention it is preferred that only certain types of the isoenzymes are reduced or inactivated while other types remain unaffected (see below). Preferably, the gene is inactivated by deletion of at least

part of the gene or by disruption of the gene, whereby in this context the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of glycerolphosphate dehydrogenase activity in the host cell.

5 A preferred gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is the *S. cerevisiae GPD2* gene as described by Eriksson et al. (1995, Mol. Microbiol. 17: 95-107), encoding the amino acid sequence of SEQ ID NO: 21 and orthologues thereof in other species. Therefore a gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or
10 inactivated in the cell of the invention preferably is a gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 21 or an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 21.

15 In a preferred embodiment of the invention, the host cell of the invention comprises a functional high-osmolarity glycerol response pathway. Preferably therefore, only the activity of the gene(s) encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 21 are reduced or inactivated, while at least one endogenous gene encoding a
20 glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 22 is functional. SEQ ID NO: 22 depicts the amino acid sequence encoded by the *S. cerevisiae GPD1* gene as described by Albertyn et al. (1994, Mol. Cell. Biol. 14: 4135-4144), which has 69% amino acid identity with the *S. cerevisiae GPD2* glycerolphosphate dehydrogenase.
25 The *S. cerevisiae GPD1* gene is the stress-induced glycerolphosphate dehydrogenase of *S. cerevisiae*, which is important for growth under osmotic stress as may occur under industrial fermentations conditions. Its expression is *inter alia* regulated by the high-osmolarity glycerol response pathway. It is therefore advantageous that a host cell of the invention has at least one functional copy of a endogenous gene encoding a
30 glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 22 or an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 22.

In a further embodiment, the activity of all the genes in the host cell encoding a glycerolphosphate dehydrogenase is reduced or inactivated. In such cells preferably all copies of endogenous genes encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to
5 SEQ ID NO: 21 or 22 (or having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 21 and 22) are inactivated or at least reduced in expression.

In a further preferred embodiment, the host cell of the invention has at least one of: a) the ability of isomerising xylose to xylulose; and, b) the ability to convert L-
10 arabinose into D-xylulose 5-phosphate. For a) the cell preferably has a functional exogenous xylose isomerase gene, which gene confers to the cell the ability to isomerise xylose into xylulose. For b) the cell preferably has functional exogenous genes coding for a L-arabinose isomerase, a L-ribulokinase and a L-ribulose-5-phosphate 4-epimerase, which genes together confers to the cell the ability to isomerise
15 convert L-arabinose into D-xylulose 5-phosphate.

Fungal host cells having the ability of isomerising xylose to xylulose as e.g. described in WO 03/0624430 and in WO 06/009434. The ability of isomerising xylose to xylulose is preferably conferred to the cell by transformation with a nucleic acid construct comprising a nucleotide sequence encoding a xylose isomerase. Preferably
20 the cell thus acquires the ability to directly isomerise xylose into xylulose. More preferably the cell thus acquires the ability to grow aerobically and/or anaerobically on xylose as sole energy and/or carbon source though direct isomerisation of xylose into xylulose (and further metabolism of xylulose). It is herein understood that the direct isomerisation of xylose into xylulose occurs in a single reaction catalysed by a xylose
25 isomerase, as opposed to the two step conversion of xylose into xylulose via a xylitol intermediate as catalysed by xylose reductase and xylitol dehydrogenase, respectively.

Several xylose isomerases (and their amino acid and coding nucleotide sequences) that may be successfully used to confer to the cell of the invention the ability to directly isomerise xylose into xylulose have been described in the art. These
30 include the xylose isomerases of *Piromyces* sp. and of other anaerobic fungi that belongs to the families *Neocallimastix*, *Caecomyces*, *Piromyces* or *Ruminomyces* (WO 03/0624430), *Cyllamyces aberensis* (US 20060234364), *Orpinomyces* (Madhavan et al., 2008, DOI 10.1007/s00253-008-1794-6), the xylose isomerase of the bacterial

genus *Bacteroides*, including e.g. *B. thetaiotaomicron* (WO 06/009434), *B. fragilis*, and *B. uniformis* (WO 09/109633), the xylose isomerase of the anaerobic bacterium *Clostridium phytofermentans* (Brat et al., 2009, Appl. Environ. Microbiol. 75: 2304–2311), and the xylose isomerases of *Clostridium difficile*, *Ciona intestinales* and
5 *Fusobacterium mortiferum* (WO 10/074577).

Fungal host cells having the ability to convert L-arabinose into D-xylulose 5-phosphate as e.g. described in Wisselink et al. (2007, AEM Accepts, published online ahead of print on 1 June 2007; Appl. Environ. Microbiol. doi:10.1128/AEM.00177-07) and in EP 1 499 708. The ability of to converting L-arabinose into D-xylulose 5-phosphate is preferably conferred to the cell by transformation with a nucleic acid
10 construct(s) comprising nucleotide sequences encoding a) an arabinose isomerase; b) a ribulokinase, preferably a L-ribulokinase a xylose isomerase; and c) a ribulose-5-P-4-epimerase, preferably a L-ribulose-5-P-4-epimerase. Preferably, in the cells of the invention, the ability to convert L-arabinose into D-xylulose 5-phosphate is the ability
15 to convert L-arabinose into D-xylulose 5-phosphate through the subsequent reactions of 1) isomerisation of arabinose into ribulose; 2) phosphorylation of ribulose to ribulose 5-phosphate; and, 3) epimerisation of ribulose 5-phosphate into D-xylulose 5-phosphate. Suitable nucleotide sequences encoding arabinose isomerases, a ribulokinases and ribulose-5-P-4-epimerases may be obtained from *Bacillus subtilis*, *Escherichia coli* (see
20 e.g. EP 1 499 708), *Lactobacilli*, e.g. *Lactobacillus plantarum* (see e.g. Wisselink et al. *supra*), or species of *Clavibacter*, *Arthrobacter* and *Gramella*, of which preferably *Clavibacter michiganensis*, *Arthrobacter aurescens* and *Gramella forsetii* (see WO2009/011591).

The transformed host cell of the invention further preferably comprises xylulose
25 kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. Preferably, the cell contains endogenous xylulose kinase activity. More preferably, a cell of the invention comprises a genetic modification that increases the specific xylulose kinase activity. Preferably the genetic modification causes overexpression of a xylulose kinase, e.g. by overexpression of a nucleotide sequence
30 encoding a xylulose kinase. The gene encoding the xylulose kinase may be endogenous to the cell or may be a xylulose kinase that is heterologous to the cell. A nucleotide sequence that may be used for overexpression of xylulose kinase in the cells of the invention is e.g. the xylulose kinase gene from *S. cerevisiae* (*XKSI*) as described by

Deng and Ho (1990, Appl. Biochem. Biotechnol. 24-25: 193-199). Another preferred xylulose kinase is a xylose kinase that is related to the xylulose kinase from *Piromyces* (*xylB*; see WO 03/0624430). This *Piromyces* xylulose kinase is actually more related to prokaryotic kinase than to all of the known eukaryotic kinases such as the yeast kinase.

5 The eukaryotic xylulose kinases have been indicated as non-specific sugar kinases, which have a broad substrate range that includes xylulose. In contrast, the prokaryotic xylulose kinases, to which the *Piromyces* kinase is most closely related, have been indicated to be more specific kinases for xylulose, i.e. having a narrower substrate range. In the cells of the invention, a xylulose kinase to be overexpressed is
10 overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity, the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme.

15 A cell of the invention further preferably comprises a genetic modification that increases the flux of the pentose phosphate pathway as described in WO 06/009434. In particular, the genetic modification causes an increased flux of the non-oxidative part pentose phosphate pathway. A genetic modification that causes an increased flux of the non-oxidative part of the pentose phosphate pathway is herein understood to mean a
20 modification that increases the flux by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to the flux in a strain which is genetically identical except for the genetic modification causing the increased flux. The flux of the non-oxidative part of the pentose phosphate pathway may be measured as described in WO 06/009434.

Genetic modifications that increase the flux of the pentose phosphate pathway
25 may be introduced in the cells of the invention in various ways. These including e.g. achieving higher steady state activity levels of xylulose kinase and/or one or more of the enzymes of the non-oxidative part pentose phosphate pathway and/or a reduced steady state level of unspecific aldose reductase activity. These changes in steady state activity levels may be effected by selection of mutants (spontaneous or induced by
30 chemicals or radiation) and/or by recombinant DNA technology e.g. by overexpression or inactivation, respectively, of genes encoding the enzymes or factors regulating these genes.

In a preferred cell of the invention, the genetic modification comprises overexpression of at least one enzyme of the (non-oxidative part) pentose phosphate pathway. Preferably the enzyme is selected from the group consisting of the enzymes encoding for ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase. Various combinations of enzymes of the (non-oxidative part) pentose phosphate pathway may be overexpressed. E.g. the enzymes that are overexpressed may be at least the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate 3-epimerase; or at least the enzymes ribulose-5-phosphate isomerase and transketolase; or at least the enzymes ribulose-5-phosphate isomerase and transaldolase; or at least the enzymes ribulose-5-phosphate 3-epimerase and transketolase; or at least the enzymes ribulose-5-phosphate 3-epimerase and transaldolase; or at least the enzymes transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate 3-epimerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, and transketolase. In one embodiment of the invention each of the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase are overexpressed in the cell of the invention. Preferred is a cell in which the genetic modification comprises at least overexpression of the enzyme transaldolase. More preferred is a cell in which the genetic modification comprises at least overexpression of both the enzymes transketolase and transaldolase as such a host cell is already capable of anaerobic growth on xylose. In fact, under some conditions we have found that cells overexpressing only the transketolase and the transaldolase already have the same anaerobic growth rate on xylose as do cells that overexpress all four of the enzymes, i.e. the ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase. Moreover, cells of the invention overexpressing both of the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate 3-epimerase are preferred over cells overexpressing only the isomerase or only the 3-epimerase as overexpression of only one of these enzymes may produce metabolic imbalances.

There are various means available in the art for overexpression of enzymes in the host cells of the invention. In particular, an enzyme may be overexpressed by

increasing the copynumber of the gene coding for the enzyme in the cell, e.g. by integrating additional copies of the gene in the cell's genome, by expressing the gene from an episomal multicopy expression vector or by introducing a episomal expression vector that comprises multiple copies of the gene. The coding sequence used for
5 overexpression of the enzymes preferably is homologous to the host cell of the invention. However, coding sequences that are heterologous to the host cell of the invention may likewise be applied. Alternatively overexpression of enzymes in the cells of the invention may be achieved by using a promoter that is not native to the sequence coding for the enzyme to be overexpressed, i.e. a promoter that is
10 heterologous to the coding sequence to which it is operably linked. Although the promoter preferably is heterologous to the coding sequence to which it is operably linked, it is also preferred that the promoter is homologous, i.e. endogenous to the cell of the invention. Preferably the heterologous promoter is capable of producing a higher steady state level of the transcript comprising the coding sequence (or is capable of
15 producing more transcript molecules, i.e. mRNA molecules, per unit of time) than is the promoter that is native to the coding sequence, preferably under conditions where one or more of xylose, arabinose and glucose are available as carbon sources, more preferably as major carbon sources (i.e. more than 50% of the available carbon source consists of one or more of xylose, arabinose and glucose), most preferably as sole
20 carbon sources.

A further preferred cell of the invention comprises a genetic modification that reduces unspecific aldose reductase activity in the cell. Preferably, unspecific aldose reductase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an unspecific aldose reductase.
25 Preferably, the genetic modifications reduce or inactivate the expression of each endogenous copy of a gene encoding an unspecific aldose reductase that is capable of reducing an aldopentose, including, xylose, xylulose and arabinose, in the cell's genome. A given cell may comprise multiple copies of genes encoding unspecific aldose reductases as a result of di-, poly- or aneu-ploidy, and/or a cell may contain
30 several different (iso)enzymes with aldose reductase activity that differ in amino acid sequence and that are each encoded by a different gene. Also in such instances preferably the expression of each gene that encodes an unspecific aldose reductase is reduced or inactivated. Preferably, the gene is inactivated by deletion of at least part of

the gene or by disruption of the gene, whereby in this context the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of unspecific aldose reductase activity in the host cell. A nucleotide sequence encoding an
5 aldose reductase whose activity is to be reduced in the cell of the invention and amino acid sequences of such aldose reductases are described in WO 06/009434 and include e.g. the (unspecific) aldose reductase genes of *S. cerevisiae* *GRE3* gene (Träff et al., 2001, Appl. Environm. Microbiol. 67: 5668-5674) and orthologues thereof in other species.

10 A further preferred transformed host cell according to the invention may comprises further genetic modifications that result in one or more of the characteristics selected from the group consisting of (a) increased transport of xylose and/or arabinose into the cell; (b) decreased sensitivity to catabolite repression; (c) increased tolerance to ethanol, osmolarity or organic acids; and, (d) reduced production of by-products. By-
15 products are understood to mean carbon-containing molecules other than the desired fermentation product and include e.g. xylitol, arabinitol, glycerol and/or acetic acid. Any genetic modification described herein may be introduced by classical mutagenesis and screening and/or selection for the desired mutant, or simply by screening and/or selection for the spontaneous mutants with the desired characteristics. Alternatively, the
20 genetic modifications may consist of overexpression of endogenous genes and/or the inactivation of endogenous genes. Genes the overexpression of which is desired for increased transport of arabinose and/or xylose into the cell are preferably chosen from genes encoding a hexose or pentose transporter. In *S. cerevisiae* and other yeasts these genes include *HXT1*, *HXT2*, *HXT3*, *HXT4*, *HXT5*, *HXT7* and *GAL2*, of which *HXT7*,
25 *HXT5* and *GAL2* are most preferred (see Sedlack and Ho, Yeast 2004; 21: 671–684). Another preferred transporter for expression in yeast is the glucose transporter encoded by the *P. stipitis* *SUT1* gene (Katahira et al., 2008, Enzyme Microb. Technol. 43: 115-119). Similarly orthologues of these transporter genes in other species may be overexpressed. Other genes that may be overexpressed in the cells of the invention
30 include genes coding for glycolytic enzymes and/or ethanologenic enzymes such as alcohol dehydrogenases. Preferred endogenous genes for inactivation include hexose kinase genes e.g. the *S. cerevisiae* *HXK2* gene (see Diderich et al., 2001, Appl. Environ. Microbiol. 67: 1587-1593); the *S. cerevisiae* *MIG1* or *MIG2* genes; genes

coding for enzymes involved in glycerol metabolism such as the *S. cerevisiae* glycerol-phosphate dehydrogenase 1 and/or 2 genes; or (hybridising) orthologues of these genes in other species. Other preferred further modifications of host cells for xylose fermentation are described in van Maris et al. (2006, Antonie van Leeuwenhoek
5 90:391–418), WO2006/009434, WO2005/023998, WO2005/111214, and WO2005/091733. Any of the genetic modifications of the cells of the invention as described herein are, in as far as possible, preferably introduced or modified by self cloning genetic modification.

A preferred host cell according to the invention has the ability to grow on at least
10 one of xylose and arabinose as carbon/energy source, preferably as sole carbon/energy source, and preferably under anaerobic conditions, i.e. conditions as defined herein below for anaerobic fermentation process. Preferably, when grown on xylose as carbon/energy source the host cell produces essentially no xylitol, e.g. the xylitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3% of the carbon
15 consumed on a molar basis. Preferably, when grown on arabinose as carbon/energy source, the cell produces essentially no arabinitol, e.g. the arabinitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3 % of the carbon consumed on a molar basis.

A preferred host cell of the invention has the ability to grow on at least one of a
20 hexose, a pentose, glycerol, acetic acid and combinations thereof at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0.25 or 0.3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. Therefore, preferably the host cell has the ability to grow on at least one of xylose and arabinose as sole carbon/energy source at a rate of at least 0.01,
25 0.02, 0.05, 0.1, 0.2, 0.25 or 0.3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. More preferably, the host cell has the ability to grow on a mixture of a hexose (e.g. glucose) and at least one of xylose and arabinose (in a 1:1 weight ratio) as sole carbon/energy source at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0.25 or 0.3 h⁻¹
30 under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. Most preferably, the host cell has the ability to grow on a mixture of a hexose (e.g. glucose), at least one of xylose and arabinose and glycerol (in a 1:1:1 weight ratio) as sole carbon/energy source

at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0.25 or 0.3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions.

In a one aspect, the invention relates to the use of a yeast cell according to the
5 invention for the preparation of a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols and isoprenoid-derived products.

In another aspect the invention relates to a process for producing a fermentation
product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic
10 acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products. The process can be a process wherein formate is produced in addition to the fermentation product. The process preferably comprises the step of: a) fermenting a medium with a yeast cell, whereby the medium contains or is fed with: a) a source of at least one of a hexose, a pentose and glycerol and whereby the yeast cell
15 ferments the at least one of a hexose, pentose and glycerol to the fermentation product, and optionally to formate. The yeast cell preferably is a (host) cell as herein defined above. The process preferably comprise one or more further steps wherein the fermentation product and/or the formate are recovered. The process may be a batch process, a fed-batch process or a continuous process as are well known in the art.

20 In a preferred process the source of at least one of a hexose, a pentose and glycerol comprises or consist of: hexose and pentose; hexose and glycerol; pentose and glycerol; hexose, pentose and glycerol. In a further preferred process, the source of at least one of a hexose, a pentose and glycerol further comprises acetic acid.

In a preferred process the source of hexose comprises or consists of glucose.
25 Preferably the source of pentose comprises or consists of at least one of xylose and arabinose, of which xylose is preferred. Preferably, the medium fermented by the cells of the invention comprises or is fed with (fractions of) hydrolysed biomass comprising at least one at least one of a hexose and a pentose such as glucose, xylose and/or arabinose. The (fractions of) hydrolysed biomass comprising the hexoses and pentose
30 will usually also comprise acetic acid (or a salt thereof). An example of hydrolysed biomass to be fermented in the processes of the invention is e.g. hydrolysed lignocellulosic biomass. Lignocellulosic biomass is herein understood as plant biomass that is composed of cellulose, hemicellulose, and lignin. The carbohydrate polymers

(cellulose and hemicelluloses) are tightly bound to the lignin. Examples of lignocellulosic biomass to be hydrolysed for use in the present invention include agricultural residues (including e.g. empty fruit bunches (EFB) of oil palm, corn stover and sugarcane bagasse), wood residues (including sawmill and paper mill discards and
5 (municipal) paper waste. Methods for hydrolysis of biomass such as lignocelluloses are known in the art per se and include e.g. acids, such as sulphuric acid and enzymes such as cellulases and hemicellulases.

In the process of the invention, the sources of xylose, glucose and arabinose may be xylose, glucose and arabinose as such (i.e. as monomeric sugars) or they may be in
10 the form of any carbohydrate oligo- or polymer comprising xylose, glucose and/or arabinose units, such as e.g. lignocellulose, arabinans, xylans, cellulose, starch and the like. For release of xylose, glucose and/or arabinose units from such carbohydrates, appropriate carbohydrases (such as arabinases, xylanases, glucanases, amylases, cellulases, glucanases and the like) may be added to the fermentation medium or may
15 be produced by the modified host cell. In the latter case the modified host cell may be genetically engineered to produce and excrete such carbohydrases. An additional advantage of using oligo- or polymeric sources of glucose is that it enables to maintain a low(er) concentration of free glucose during the fermentation, e.g. by using rate-limiting amounts of the carbohydrases preferably during the fermentation. This, in turn,
20 will prevent repression of systems required for metabolism and transport of non-glucose sugars such as xylose and arabinose. In a preferred process the modified host cell ferments both the glucose and at least one of xylose and arabinose, preferably simultaneously in which case preferably a modified host cell is used which is insensitive to glucose repression to prevent diauxic growth. In addition to a source of at
25 least one of a hexose, a pentose and glycerol, as carbon source, the fermentation medium will further comprise the appropriate ingredients required for growth of the modified host cell. Compositions of fermentation media for growth of eukaryotic microorganisms such as yeasts are well known in the art.

In the process of the invention, the medium further preferably comprises and/or is
30 fed a source of glycerol. Glycerol for use in the process of the present invention may advantageously be glycerol that is generated as a by-product in biodiesel production from transesterification reactions using vegetable oils, such as palm oil, or animal fats and an alcohol.

The fermentation process may be an aerobic or an anaerobic fermentation process. An anaerobic fermentation process is herein defined as a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than 5, 2.5 or 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both electron donor and electron acceptors. In the absence of oxygen, NADH produced in glycolysis and biomass formation, cannot be oxidised by oxidative phosphorylation. To solve this problem many microorganisms use pyruvate or one of its derivatives as an electron and hydrogen acceptor thereby regenerating NAD^+ . Thus, in a preferred anaerobic fermentation process pyruvate is used as an electron (and hydrogen acceptor) and is reduced to fermentation products such as ethanol, as well as non-ethanol fermentation products such as lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products, preferably under concomitant production of formate. Anaerobic processes of the invention are preferred over aerobic processes because anaerobic processes do not require investments and energy for aeration and in addition, anaerobic processes produce higher product yields than aerobic processes. Alternatively, the fermentation process of the invention may be run under aerobic oxygen-limited conditions. Preferably, in an aerobic process under oxygen-limited conditions, the rate of oxygen consumption is at least 5.5, more preferably at least 6 and even more preferably at least 7 mmol/L/h.

The fermentation process is preferably run at a temperature that is optimal for the modified cells of the invention. Thus, for most yeasts cells, the fermentation process is performed at a temperature which is less than 42°C, preferably less than 38°C. For yeast cells, the fermentation process is preferably performed at a temperature which is lower than 35, 33, 30 or 28°C and at a temperature which is higher than 20, 22, or 25°C.

Because undissociated formic acid is more toxic to the host cells of the invention as compared to the dissociated acid ion, the fermentation process is preferably run at a pH that is higher than the pKa of formic acid, which is 3.75. The fermentation process is therefore preferably run at a pH of at least 3.8, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5 or 8.0. Preferably the fermentation process is run at at least the aforementioned pHs for the entire duration of the process. Alternatively the pH may be regulated so as to maintain the concentration of undissociated formic acid below a given concentration.

Preferably the pH of the medium is regulated during the fermentation process to maintain a concentration of undissociated formic acid that is not higher than 50.0, 30.0, 20.0, 18.1, 15.0, 10.0, 5.0, 2.0 or 1.0 mM.

A preferred fermentation process according to the invention is a process for the
5 production of ethanol and optionally formate, whereby the process comprises the step
of fermenting a medium with a yeast cell, whereby the medium contains or is fed with a
source of at least one of a hexose, a pentose and glycerol and whereby the yeast cell
ferments the at least one of a hexose, pentose and glycerol to ethanol and optionally
formate. Optionally the process comprises the step of recovery of at least one of ethanol
10 and optionally formate. The fermentation may further be performed as described above.
In the process the volumetric ethanol productivity is preferably at least 0.5, 1.0, 1.5,
2.0, 2.5, 3.0, 5.0 or 10.0 g ethanol per litre per hour. The ethanol yield on hexose and/or
pentose and/or glycerol (and/or acetate) in the process preferably is at least 50, 60, 70,
80, 90, 95 or 98%. The ethanol yield is herein defined as a percentage of the theoretical
15 maximum yield, which, for xylose, glucose and arabinose is 0.51 g. ethanol per g.
hexose or pentose. For glycerol the theoretical maximum yield is 0.50 g. ethanol per g.
glycerol and for acetic acid the theoretical maximum yield is 0.77 g. ethanol per g.
acetic acid.

In this document and in its claims, the verb "to comprise" and its conjugations is
20 used in its non-limiting sense to mean that items following the word are included, but
items not specifically mentioned are not excluded. In addition, reference to an element
by the indefinite article "a" or "an" does not exclude the possibility that more than one
of the element is present, unless the context clearly requires that there be one and only
one of the elements. The indefinite article "a" or "an" thus usually means "at least one".

25 The following examples are offered for illustrative purposes only, and are not
intended to limit the scope of the present invention in any way.

Examples

1. Enzyme activity assays

Cell extracts for activity assays were prepared from exponentially growing aerobic or anaerobic batch cultures and analysed for protein content as described by
5 Abbot et al., (2009, Appl. Environ. Microbiol. 75: 2320-2325).

NAD⁺-dependent acetaldehyde dehydrogenase (EC 1.2.1.10) activity was measured at 30 °C by monitoring the oxidation of NADH at 340 nm. The reaction mixture (total volume 1 ml) contained 50 mM potassium phosphate buffer (pH 7.5), 0.15 mM NADH and cell extract. The reaction was started by addition of 0.5 mM
10 acetyl-Coenzyme A.

For glycerol 3-phosphate dehydrogenase (EC 1.1.1.8) activity determination, cell extracts were prepared as described above except that the phosphate buffer was replaced by triethanolamine buffer (10 mM, pH 5). Glycerol-3-phosphate dehydrogenase activities were assayed in cell extracts at 30 °C as described previously
15 (Blomberg and Adler, 1989, J. Bacteriol. 171: 1087-1092.9). Reaction rates were proportional to the amounts of cell extract added.

Acetyl-CoA synthase (EC 6.2.1.1) activity was measured as described by Frenkel and Kitchens (1977, J. Biol. Chem. 252: 504-507) which is a modification of the method of Webster (Webster, 1969, Methods Enzymol. 13: 375-381). NADH
20 formation measured is spectrophotometrically when the acetyl-CoA produced is coupled with citrate synthase and malate dehydrogenase reactions. The assay system contained 100 mM Tris-Cl(pH 7.6), 10 mM MgCl₂, 6 mM ATP, 5 mM malate, 1 mM NAD⁺, 0.1 mM NADH, 2.5 mM dithiothreitol or 2-mercaptoethanol, 0.2 mM coenzyme A, 25 µg citrate synthase (80 units/mg), 20 µg malate dehydrogenase (1000
25 units/mg), and 10 mM acetate and the reaction was measured rate was measured at 340 nm and calculated from the extinction coefficient of NADH ($6.22 \times 10^6 \text{ cm}^2/\text{mol}$).

The activity of glycerol dehydrogenase and dihydroxyacetone kinase are measured at 30 °C in cell extracts, essentially as previously described (Gonzalez et al., 2008, Metab. Eng. 10, 234–245). Enzyme activities of glycerol dehydrogenase and
30 dihydroxyacetone kinase are reported as µmoles of substrate/min/mg of cell protein.

2. Strain construction

All modifications start with the xylose fermenting strain RN1041. RN1041 is a CEN.PK based strain with the following genotype:

Mat a, ura3-52, leu2-112, his3::loxP, gre3::loxP, loxP-Ptpi::TAL1, loxP-Ptpi::RKII, loxP-Ptpi-TKL1, loxP-Ptpi-RPE1, delta::Padh1XKS1Tcyc1-LEU2, delta::URA3-Ptpi-xylA-Tcyc1

Mat a = mating type a

ura3-52, leu2-112, his3::loxP mutations in the genes *ura3*, *leu2* and *his3*, the *ura3* is complemented by the *Piromyces xylA* overexpression construct, *leu2* is complemented by the XKS1 overexpression construct. *his3* could be used for selection of additional plasmids, RN1041 needs histidine in the medium for growth.

gre3::loxP = deletion of the *gre3* gene encoding xylose reductase, *loxP* site is left after marker removal.

loxP-Ptpi= overexpression of het pentose phosphate pathway, *loxP* site upstream of constitutive promoter is left after marker removal

delta:: = integration of the construct after recombination on the long terminal repeats of the Ty1 retrotransposon.

Constructs for expression of the *E.coli* pflA and pflB genes in yeast

The *E.coli* pflA and pflB genes were obtained as codon optimized synthetic genes.

The pflA expression construct by ligating a TEF1 promoter-fragment (cut with the restriction enzymes *SalI* and *HindIII*; from plasmid collection Royal Nidalco), a pflA synthetic ORF (cut with *HindIII* and *BssHII*) and an ADH1 terminator-fragment (cut with *BssHII* and *BsiWI*; from plasmid collection Royal Nidalco) together into pCRII blunt (Invitrogen) to yield pRN613.

The pflB expression construct by ligating a PGK1 promoter-fragment (cut with the restriction enzymes *SpeI* and *PstI*; from plasmid collection Royal Nidalco), a pflB synthetic ORF (cut with *PstI* and *SalI*) and an PGI1 terminator-fragment (cut with *XhoI* and *BsiWI*; from plasmid collection Royal Nidalco) together into pCRII blunt (Invitrogen) to yield pRN614.

The LEU2 marker from plasmid pRS305 (Sikorski and Hieter, 1989, Genetics 122:19-27) is exchanged for the ZeoMX marker (*zeoMX* = resistance to phleomycin due to expression of the Tn5 ble gene with a TEF1 promoter and terminator sequence of

Ashbya gossypii). Yeast transformants are selected on media containing Zeomycin. pRS305 is cut with *Bst*BI and *Bsr*GI and a ZeoMX fragment (from plasmid collection Royal Nedalco) is cut with the same restriction enzymes. Ligation gives plasmid pRS30z.

- 5 A 2 μ ORI is introduced by cutting pRS30z with *Afl*III en *Mfe*I and cutting a 2 μ fragment (from plasmid collection Royal Nedalco) with *Afl*III and *Eco*RI. The fragments are ligated to yield pRN615 (6341bp)

For the final construction pRN615 is cut with *Xho*I and *Spe*I, this vector is combined with the inserts from pRN613 cut with *Sal*I and *Bsi*WI (1510bp) and pRN614cut with
10 *Spe*I and *Bsi*WI (3381bp) to produce pRN616 (SEQ ID NO: 31).

For expression of the *E.coli* pflA gene only, pRN613 is cut with *Sal*I and *Bsi*WI and this fragment is ligated into pRN615 cut with *Acc*65I and *Xho*I to produce pRN619 (SEQ ID NO: 32).

For expression of the *E.coli* pflB gene only, pRN614 is cut with *Xho*I and *Bsi*WI and
15 this fragment is ligated into pRN615 cut with *Acc*65I and *Xho*I to produce pRN620 (SEQ ID NO: 33).

Plasmid pRN616, pRN619 and pRN620 are used to transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

20

Constructs for overexpression of GCY1, gldA, DAK1 dhaK, and GUP1

GCY1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Pst*I site 5'of the ATG and an *Afl*III site 3'of the TAA to produce the fragment of SEQ ID
25 NO: 25. A DNA fragment comprising the *S. cerevisiae* Actin1 promoter is ligated upstream of the GCY1 ORF and DNA fragment comprising the *S. cerevisiae* ADH1 terminator fragment is ligated downstream of the GCY1 ORF.

gldA

The construct for expression in yeast of the *E.coli* gldA was made by ligating a yeast
30 ACT1 promoter fragment (cut with the restriction enzymes *Spe*I and *Pst*I), a synthetic ORF (SEQ ID NO:50), encoding the *E.coli* gldA, (cut with *Pst*I en *Bss*HII) and a yeast CYC1 terminator fragment (cut with *Bss*HII and *Bsi*WI) together into pCRII blunt (Invitrogen) to yield pRNgldA (SEQ ID NO:51).

DAK1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Xba*I site 5' of the ATG and a *Sal*I site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 26. A DNA fragment comprising the *S. cerevisiae* TPI1 promoter is
 5 ligated upstream of the DAK1 ORF and DNA fragment comprising the *S. cerevisiae* PGI1 terminator fragment is ligated downstream of the DAK1 ORF to produce the expression construct of SEQ ID NO: 26.

dhaK

The construct for expression in yeast of the *Citrobacter freundii* *dhaK* was made by
 10 ligating the yeast TPI1 promoter fragment (cut with the restriction enzymes *Xho*I and *Xba*I), a synthetic ORF (SEQ ID NO:53), encoding the *C. freundii* *dhaK*, (cut with *Xba*I and *Sal*I) and a yeast PGI1 terminator fragment (cut with *Xho*I and *Bsi*WI) together into pCRII blunt (Invitrogen) to yield pRNdhaK (SEQ ID NO:54).

GUP1

15 PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Hind*III site 5' of the ATG and a *Bam*HI site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 27. A DNA fragment comprising the *S. cerevisiae* TDH3 promoter is ligated upstream of the GUP1 ORF and DNA fragment comprising the *S. cerevisiae* CYC1 terminator fragment is ligated downstream of the GUP1 ORF to produce the
 20 expression construct of SEQ ID NO: 27.

Strains overexpressing GCY1, DAK1 and GUP1

The above-described expression constructs of GCY1, DAK1 and GUP1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN605 (SEQ ID NO: 29). Plasmid pRN605 is used to
 25 transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

The above-described expression constructs of GCY1 and DAK1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN608 (SEQ ID NO: 30). is used to transform the yeast
 30 strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strain RN1098, as indicated in Table 7.

Strains expressing *E.coli* gldA with or without expression of *C.freundii* dhaK

The *E.coli* gldA expression construct is cut from plasmid pRN_{gldA} with the restriction enzymes *SpeI* and *BsiWI*. The *C.freundii* dhaK expression construct is cut from plasmid pRN_{dhaK} with the restriction enzymes *BsiWI* and *XhoI*. These fragments are ligated into plasmid pRN595 cut with the restriction enzymes *SpeI* and *SalI* to yield
 5 pRN957 (SEQ ID NO: 55). Plasmid pRN958 (SEQ ID NO: 56) is subsequently made from plasmid pRN957 by deleting the *C.freundii* dhaK expression construct. To this end plasmid pRN957 is cut with the restriction enzymes *PspOMI* and *AvrII*, the single stranded parts are filled in with the DNA polymerase phusion (Finnzymes) and the plasmid is religated.

10 The expression constructs of DAK1 and GUP1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN607 (SEQ ID NO: 57).

Plasmids pRN957, pRN958 and pRN607 used to transform yeast together with the other plasmids indicated in Table 7 to produce strain RN1194, RN1195 and RN1196.

15

Construct for overexpression of FPS1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *NsiI* site 5' of the ATG and a *BamHI* site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 28. A DNA fragment comprising the *S. cerevisiae* ADH1(medium)
 20 promoter is ligated upstream of the FSP1 ORF and DNA fragment comprising the *S. cerevisiae* CYC1 terminator fragment is ligated downstream of the FSP1 ORF to produce the expression construct of SEQ ID NO: 28. Ligation of these fragments in pCRII blunt produces pRN617. The insert of pRN617 is cut with *BsiWI* and *XhoI* and cloned in pRN616 cut with *BsrGI* and *SalI* to produce pRN618 (SEQ ID NO: 34).
 25 Plasmid pRN618 is used to transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

Constructs for deletion of FDH1 and FDH2

Primers FDHuf, FDHur, FDHdf and FDHdr are used for amplification of genomic sequences fragments upstream and downstream of both FDH1 and FDH2 genes for
 30 their inactivation. The same primers are used for both FDH1 and FDH2 alleles.

For inactivation of the FDH1 allele, a 423bp upstream PCR fragment with a *BspEI* site at the 3'-end for ligation to the MX marker, and a *BbsI* site at the 5'-end for isolation of the deletion construct is amplified using FDHuf and FDHur and cloned in pCRII

(blunt) using topo-cloning (Invitrogen), which results in an additional *Bst*BI site at the 5'-end.

FDHuf: TCGAAGACTCCGAATGAAAAAGACATGCCAG (SEQ ID NO: 35)

FDHur: TCCGGATACCAAGTTCATTTTCAATACACCCCA (SEQ ID NO: 36)

5

A 378bp downstream PCR fragment with a *Nsi*I and a *Sph*I site at the 5'-end for ligation to the MX marker, and a *Bbs*I site at the 3'-end for isolation of the deletion construct is amplified using FDHdf and FDHdr.

10 FDHdf: ATGCATGCAGAATGGTTCTTATGCCAC (SEQ ID NO: 37)

FDHdr: GAAGACAGTTCTGTTATTAACGACGAGCCA (SEQ ID NO: 38)

For the final construction the downstream fragment cut with *Xho*I and *Nsi*I is ligated to the patMX marker (collection Royal Nedalco) cut with *Bsp*EI and *Nsi*I in the plasmid
15 containing the upstream-fragment cut with *Bsp*EI and *Xho*I. The final plasmid, pRN621 (SEQ ID NO: 39), is cut with *Bbs*I prior to yeast transformation. Transformants are selected for phosphinotricin resistance. Correct integration of the disruption fragment is verified by PCR with specific primers.

FDH2 is found as a pseudogene in some strains (e.g. S288C, YPL275W and
20 YPL276W). Two mutations restore the ORF homology with FDH1 (T436C and del A476). Sequence analysis shows that in the CEN.PK background (RN1041) FDH2 is one ORF homologous to FDH1.

For the final construction the downstream fragment cut with *Xho*I and *Sph*I is ligated to the kanMX marker (collection Royal Nedalco) cut with *Bsp*EI and *Sph*I into the
25 plasmid containing the upstream fragment cut with *Bsp*EI and *Xho*I. The final plasmid, pRN622 (SEQ ID NO: 40), is cut with *Bbs*I prior to yeast transformation. Transformants are selected for G418 resistance. Correct integration of the disruption fragment is verified by PCR with specific primers.

Disruption fragments from pRN621 and pRN622 are used to transform the yeast strain
30 RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

Constructs for expression of the *E.coli* mhpF, adhE or the *Entamoeba histolytica* adh2 genes in yeast

For expression of the *E.coli* mhpF gene, a yeast PGK1 promoter fragment (*SpeI-PstI*) and an ADH1 terminator fragment (*AflIII-NotI*) (both from the Nedalco plasmid collection) were ligated onto the codon-optimized synthetic fragment encoding the *E.coli* mhpF (SEQ ID NO: 8).

- 5 pRS 303 with 2 μ ori (= pRN347, Royal Nedalco plasmid collection) was cut with *SpeI* and *NotI* and the mhpF expression construct was cloned into this vector to produce pRN558 (SEQ ID NO: 41). pRN558 is used to transform the yeast strain RN1041 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of mhpF is
10 verified by qPCR.

- For expression of the *E.coli* adhE gene, a codon optimized synthetic fragment encoding the *E.coli* adhE (SEQ ID NO: 10) is cut with *XbaI* and *AflIII* and ligated into pRN558 cut with *XbaI* and *AflIII* (replacing the *E.coli* mhpF gene in pRN558) to produce pRN595 (SEQ ID NO: 42). pRN595 is used to transform the yeast strain RN1041
15 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of adhE is verified by qPCR.

- For expression of the *Entamoeba histolytica* adh2, a codon optimized synthetic fragment encoding the *E. histolytica* adh2 (SEQ ID NO: 12) is cut with *XbaI* and *AflIII*
20 and ligated into pRN558 cut with *XbaI* and *AflIII* (replacing the *E.coli* mhpF gene in pRN558) to produce pRN596 (SEQ ID NO: 43). pRN596 is used to transform the yeast strain RN1041 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of adh2 is verified by qPCR.

25 Constructs for deletion of GPD2

- Primers GPD2uf, GPD2ur, GPD2df and GPD2dr are used for amplification of genomic sequences fragments upstream and downstream of the GPD2 gene for its inactivation. A 407bp upstream PCR fragment with an *AflIII* site at the 3'-end (derived from the GPD2 sequence) and a *BglIII* site at the 5'-end (for isolation of the deletion construct) is
30 amplified using GPD2uf, GPD2ur and cloned in pCR2.1 (topo T/A, Invitrogen).

GPD2uf: GGTACCAGATCTTTTGCGGCGAGGTGCCG (SEQ ID NO: 44)

GPD2_{ur}: TCTAGACTTAAGGAATGTGTATCTTGTTAATCTTCTGACAGC (SEQ ID NO: 45)

A 417bp downstream PCR fragment with a *Xho*I site at the 5'-end and a *Bgl*II site at the 3'-end is amplified using GPD2_{df} and GPD2_{dr}.

GPD2_{df}: CTCGAGATAGTCTACAACAACGTCCGCA (SEQ ID NO: 46)

GPD2_{dr}: CCATGGAGATCTGCAGTGAAAAAGCTCGAAGAAACAGCT (SEQ ID NO: 47)

10

For the final construction the plasmid containing the upstream fragment is cut with *Afl*III and *Kpn*I, the downstream fragment is cut with *Xho*I and *Nco*I and the natMX marker (plasmid collection Royal Nidalco) is cut with *Afl*III and *Xho*I and the fragments are ligated to produce plasmid pRN594 (SEQ ID NO: 48). pRN594 is cut with *Bgl*II prior to yeast transformation. pRN594 is used to transform the yeast strain RN1041 (selection for nourseotricin resistance) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Correct integration is verified by PCR using specific primers.

20 3. Fermentations with the constructed strains

3.1 Experimental set-up

Strains RN1041, RN1088, RN1089, RN1090, RN1091, RN1092, RN1093, RN1094, RN1095, RN1096, RN1097, RN1098, RN1099, RN1194, RN1195 and RN1196. are all tested for their ability to grow under both anoxic conditions (no oxygen present) at the expense of glycerol as carbon source. At the same time, formation of ethanol and formic acid is monitored. In several instances it is confirmed by experiments that the glycerol-consuming strains also retain their ability to ferment glucose and xylose.

Precultures of strains are prepared by inoculating a frozen glycerol stock culture of the yeast in an YP (Yeast extract Peptone) medium with addition of the sugar glucose (2.5% w/v) at 32 °C and pH 5.5. After 24 h incubation, this culture is used to inoculate the fermenter cultures. Cells are harvested by centrifugation and washed with cold dH₂O. Yeast inoculation used is 2 gram dry matter yeast per litre of fermentation medium.

Fermentations are performed as chemostat cultivation at 32 °C in 1-liter fermenters with a working volume of 800 ml. The dilution rate is set at 0.1 h⁻¹. These anoxic batch fermentations are performed in YP medium to which glycerol in combination with xylose and/or glucose is added. The culture pH is kept at pH = 7.0 by automatic
5 addition of 2M KOH. Cultures were stirred at 300 rpm.

Samples for analysis of glucose, xylose, ethanol, glycerol and formic acid production are taken at steady state situations which are obtained after 5 volume changes of the fermentation vessel. Ethanol, glycerol and formic acid concentrations are monitored by HPLC analysis. Glucose and xylose are determined by HPAEC (Dionex) analysis.

10 3.2 Results

Strain RN1041 (Tables 6 and 7) is a strain that is derived from strain RN1001 (RN1001 his3::loxP). Strain RN1001 has been described previously. If either glucose (25 mM) or xylose (25 mM) are added to the YP medium, then the anoxic fermentations under the conditions described above, leads to a full consumption of either glucose or xylose by
15 this strain. The concentrations of ethanol in the glucose and xylose media, respectively, are 40 mM and 33 mM at steady state. Small amounts of glycerol are formed both for glucose and xylose (1.8 and 1.5 mM, respectively). No formic acid formation is observed. In the absence of sugars in the YP medium, no ethanol nor glycerol or formic acid is formed. If glycerol was added to the medium at 50 mM concentration, then the
20 organism did not produce additional ethanol or formic acid as compared to the situations with glucose or xylose alone.

Strain RN1088 was derived from strain RN1041 by

1. upregulating DAK1, GCY1 and GUP1 allowing for enhanced uptake and consumption of glycerol under anoxic conditions
- 25 2. heterologous expression of pflA + pflB allowing for the conversion of pyruvate into formic acid and acetylCoA
3. heterologous expression of adhE2 allowing the conversion of acetylCoA into ethanol
4. deleting FDH1 and FDH2 preventing the oxidation of formic acid by this strain.

30 The organism is tested for growth and product formation in the same way as RN1041 as described above. With either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol was slightly lower and traces of formic acid are observed. If glycerol is added to either the

xylose- or glucose-containing medium at 50 mM concentration, then the organism does not consume more glycerol or produce additional ethanol and furthermore no additional formic acid is produced. Strains RN1089 and RN1090 have a similar genetic background as RN1088. The only difference is that adhE is not expressed in these organisms. Instead adh2 or mhpF are expressed. The characteristics of these strains during glycerol fermentations are similar to the characteristics of strain RN1088. Both strains showed similar characteristics as RN1088 for both glucose or xylose fermentation as strain RN1088.

Strains RN1091, RN1092 and RN1093 have the same genetic background as strain RN1088 with the exception of formic acid dehydrogenases. These strains have either 1 or 2 genes still present encoding for a formate dehydrogenase. Neither of the 3 organisms is able to perform good fermentations on glycerol. Only traces of either ethanol and formic acid were seen during fermentation experiments. The organisms do ferment both glucose or xylose at rates comparable to strain RN1088.

Strains RN1094, RN1095 and RN1096 have the same genetic background as strain RN1088 with the exception of the expression of either pflA or pflB. These strains lack either both these genes or either of the genes. Neither of the 3 organisms is able to perform good fermentations on glycerol. No additional ethanol or formic acid are seen as based on the addition of 50 mM glycerol to xylose-containing medium (25 mM) during fermentation experiments. The organisms do ferment both glucose or xylose at rates comparable to strain RN1088.

Strain RN1097 is derived from strain RN1088 by deleting GPD2. In doing so, a partial operation of a futile cycle might be prevented at the level of glycerol consumption/production. With glucose or xylose, the ethanol concentration at steady states are 38mM and 31 mM, respectively. Glycerol accumulates at 0.9mM and 0.5mM, respectively and formic acid had accumulates at 1.0 mM and 0.8 mM respectively. If glycerol is added to either the xylose- or glucose-containing medium at 50 mM concentration, then the organism does not consume more glycerol or produce additional ethanol and furthermore no additional formic acid is produced.

Strains RN1098 and RN1099 resemble strain RN1088 but differ genetically at the level of the glycerol transporter(s). Strain RN1098 has FPS1 overexpressed instead of GUP1, whereas strain RN1099 has both transporter genes expressed. The organisms are tested for growth and product formation in the same way as RN1041 as described above. With

either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol was slightly lower and traces of formic acid were observed. If glycerol was added to either the xylose- or glucose-containing medium at 50 mM concentration, then the RN1098 and RN1099 organisms did not
5 consume more glycerol or produce additional ethanol and furthermore no additional formic acid was produced.

RN1194 is tested for growth and product formation in the same way as RN1041 as described above. With either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol is slightly
10 lower and traces of formic acid were observed. If glycerol is added to either the xylose- or glucose-containing medium at 50 mM concentration, then the organism produces additional ethanol and furthermore formic acid is produced. The concentration of ethanol at steady state in the glycerol/glucose medium is 75 mM and the formic acid concentration is 32 mM. The remaining glycerol level is 7 mM. The concentration of
15 ethanol at steady state in the glycerol/xylose medium is 66 mM and the formic acid concentration is 30 mM. The remaining glycerol level is 8 mM. No attempts are made to test for any other fermentation products that might have been produced on the basis of glycerol. RN1196 produces similar results as RN1194 although glycerol consumption and production of ethanol and formic acid are somewhat reduced
20 compared to RN1194.

Strain RN1195 is derived from strain RN1194 by deleting GPD2. In doing so, a partial operation of a futile cycle might be prevented at the level of glycerol consumption/production. With glucose or xylose, the ethanol concentration at steady states are 38mM and 31 mM, respectively. Glycerol accumulates at 0.9mM and
25 0.5mM, respectively and formic acid accumulates at 1.0 mM and 0.8 mM, respectively. If glycerol is added to the xylose-medium (25 mM) at 50 mM concentration, then the organism additionally produces both ethanol and formic acid. The final concentration of ethanol at steady state was 72 mM, for glycerol 2 mM, and for formic acid 37 mM. No attempts were made to test for any other fermentation products that might have
30 been produced on the basis of glycerol. These results clearly indicate that deletion of the GPD2 gene has a beneficial effect on the fermentation of glycerol in strain RN1195 as compared to strain RN1194.

Table 6

Gene	GPD2	adhE	ade2	mphF	DAK1	GCY1	GUP1	FPS1	FDH1	FDH2	pflA+pflB	E.c gldA	C.f. dhak
Strain / marker		HIS3	HIS3	HIS3	hphMX	hphMX	hphMX	zeoMX	patMX	kanMX	zeoMX	HIS3	HIS3
RN1041	wt	absent	absent	absent	wt	wt	wt	wt	wt	wt	absent/absent	absent	absent
RN1088	wt	expr	absent	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1089	wt	absent	expr	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1090	wt	absent	absent	expr	up	up	up	wt	del	del	expr/expr	absent	absent
RN1091	wt	expr	absent	absent	up	up	up	wt	wt	wt	expr/expr	absent	absent
RN1092	wt	expr	absent	absent	up	up	up	wt	del	wt	expr/expr	absent	absent
RN1093	wt	expr	absent	absent	up	up	up	wt	wt	del	expr/expr	absent	absent
RN1094	wt	expr	absent	absent	up	up	up	wt	del	del	absent/absent	absent	absent
RN1095	wt	expr	absent	absent	up	up	up	wt	del	del	expr/absent	absent	absent
RN1096	wt	expr	absent	absent	up	up	up	wt	del	del	absent/expr	absent	absent
RN1097	del	expr	absent	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1098	wt	expr	absent	absent	up	up	wt	up	del	del	expr/expr	absent	absent
RN1099	wt	expr	absent	absent	up	up	up	up	del	del	expr/expr	absent	absent
RN1194	wt	expr	absent	absent	wt	wt	wt	wt	del	del	expr/expr	expr	expr
RN1195	del	expr	absent	absent	wt	wt	wt	wt	del	del	expr/expr	expr	expr
RN1196	wt	expr	absent	absent	up	wt	up	wt	del	del	expr/expr	expr	absent

Table 7

	plasmids					
selection:	HIS3	hph	loxPNatMXlox P	nat	kan	zeo
Strain						
RN1041	pRN347					
RN1088	pRN595	pRN605		pRN621	pRN622	pRN616
RN1089	pRN596	pRN605		pRN621	pRN622	pRN616
RN1090	pRN558	pRN605		pRN621	pRN622	pRN616
RN1091	pRN595	pRN605				pRN616
RN1092	pRN595	pRN605		pRN621		pRN616
RN1093	pRN595	pRN605			pRN622	pRN616
RN1094	pRN595	pRN605		pRN621	pRN622	
RN1095	pRN595	pRN605		pRN621	pRN622	pRN619
RN1096	pRN595	pRN605		pRN621	pRN622	pRN620
RN1097	pRN595	pRN605	pRN594	pRN621	pRN622	pRN616
RN1098	pRN595	pRN608		pRN621	pRN622	pRN618
RN1099	pRN595	pRN605		pRN621	pRN622	pRN618
RN1194	pRN957			pRN621	pRN622	pRN616
RN1195	pRN957		pRN594	pRN621	pRN622	pRN616
RN1196	pRN958	pRN607		pRN621	pRN622	pRN616

pRN347 = pRS303+2 μ ORI

SEQUENCE LISTING IN ELECTRONIC FORM

In accordance with Section 111(1) of the Patent Rules, this description contains a sequence listing in electronic form in ASCII text format (file: 52215-155 Seq 27-SEP-13 v1.txt).

A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

The sequences in the sequence listing in electronic form are reproduced in the following table.

SEQUENCE TABLE

<110> DSM IP Assets B.V.

<120> Yeast strains engineered to produce ethanol from glycerol

<130> 52215-155

<140> CA national phase of PCT/NL2011/050787

<141> 2011-11-18

<150> EP 10191736.7

<151> 2010-11-18

<150> US 61/415,054

<151> 2010-11-18

<150> US 61/471,836

<151> 2011-04-05

<160> 57

<170> PatentIn version 3.3

<210> 1

<211> 760

<212> PRT

<213> Escherichia coli

<400> 1

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

Gly Leu Ser Val Ala Ala Asp Ser Leu Ser Ala Ile Lys Tyr Ala Lys
 530 535 540
 Val Lys Pro Ile Arg Asp Glu Asp Gly Leu Ala Ile Asp Phe Glu Ile
 545 550 555 560
 Glu Gly Glu Tyr Pro Gln Phe Gly Asn Asn Asp Pro Arg Val Asp Asp
 565 570 575
 Leu Ala Val Asp Leu Val Glu Arg Phe Met Lys Lys Ile Gln Lys Leu
 580 585 590
 His Thr Tyr Arg Asp Ala Ile Pro Thr Gln Ser Val Leu Thr Ile Thr
 595 600 605
 Ser Asn Val Val Tyr Gly Lys Lys Thr Gly Asn Thr Pro Asp Gly Arg
 610 615 620
 Arg Ala Gly Ala Pro Phe Gly Pro Gly Ala Asn Pro Met His Gly Arg
 625 630 635 640
 Asp Gln Lys Gly Ala Val Ala Ser Leu Thr Ser Val Ala Lys Leu Pro
 645 650 655
 Phe Ala Tyr Ala Lys Asp Gly Ile Ser Tyr Thr Phe Ser Ile Val Pro
 660 665 670
 Asn Ala Leu Gly Lys Asp Asp Glu Val Arg Lys Thr Asn Leu Ala Gly
 675 680 685
 Leu Met Asp Gly Tyr Phe His His Glu Ala Ser Ile Glu Gly Gly Gln
 690 695 700
 His Leu Asn Val Asn Val Met Asn Arg Glu Met Leu Leu Asp Ala Met
 705 710 715 720
 Glu Asn Pro Glu Lys Tyr Pro Gln Leu Thr Ile Arg Val Ser Gly Tyr
 725 730 735
 Ala Val Arg Phe Asn Ser Leu Thr Lys Glu Gln Gln Gln Asp Val Ile
 740 745 750
 Thr Arg Thr Phe Thr Gln Ser Met
 755 760

<210> 2

<211> 2283

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic codon optimised E.coli pflB

<400> 2

atgtctgaat tgaacgagaa gttggctacc gcttgggaag gtttcaccaa gggtgactgg 60
 caaaacgaag ttaacgtag agacttcac caaagaact acacccata cgaaggtag 120
 gaatctttct tggctggtgc taccgaagct accaccacct tgtgggacaa gggtatggaa 180
 ggtgttaagt tggaaaacag aaccacgct ccagttgact tcgacaccgc tgttgcttct 240
 accatcacct ctacgacgc tggttacatc aacaagcaat tggaaaagat cgttggttta 300
 caaacgaag ctccattgaa gagagctttg atccattcg gtggtatcaa gatgatcgaa 360
 gggtcttgta aggcttaca cagagaattg gacccaatga tcaagaagat tttcaccgaa 420
 tacagaaaga cccacaacca aggtgttttc gacgtttaca ctccagacat cttgagatgt 480
 agaaagtctg gtgttttgac tggtttgcca gacgcttacg gtagaggtag aatcatcggt 540
 gactacagaa gagttgcttt gtacggtatc gactacttga tgaaggacaa gttggctcaa 600
 ttcacctctt tgcaagctga cttggaaaac ggtgttaact tggaaacaa catcagattg 660
 agagaagaaa tcgctgaaca acacagagct ttgggtcaaa tgaaggaaat ggctgctaag 720
 tacggttacg acatctctgg tccagctacc aacgctcaag aagctatcca atggacctac 780
 ttcggttact tggctgctgt taagtctcaa aacggtgctg ctatgtcttt cggtaggacc 840
 tctaccttct tggacgttta catcgaaaga gacttgaagg ctggtaatgat caccgaacaa 900

```

gaagctcaag aaatggttga ccacttggtt atgaagttga gaatggttag attcttgaga 960
accccagaat acgacgaatt gttctctggt gacccaatct gggctaccga atctatcggt 1020
ggtatgggtt tggacggttag aaccttggtt accaagaact ctttcagatt cttgaacacc 1080
ttatacacca tgggtccatc tccagaacca aacatgacca tcttggtggtc tgaaaagtta 1140
ccattgaact tcaagaagtt cgctgctaag gtttctatcg acacctcttc tttgcaatac 1200
gaaaacgacg acttgatgag accagacttc aacaacgacg actacgctat cgcttggtgt 1260
gtttctccaa tgatcggttg taagcaaagt caattcttcg gtgctagagc taacttggtt 1320
aagaccatgt tgtacgctat caacggtggt gttgacgaaa agttgaagat gcaagttggt 1380
ccaaagtctg aaccaatcaa gggtagcgtt ttgaactacg acgaagttat ggaaagaatg 1440
gaccacttca tggactgggt ggctaagcaa tacatcaccg ctttgaacat catccactac 1500
atgcacgaca agtactctta cgaagcatca ttgatggctt tgcacgacag agacgtaatc 1560
agaaccatgg cttgtggtat cgctgggttg tctggtgctg ctgactcttt gtctgctatc 1620
aagtacgcta aggttaagcc aatcagagac gaagacggtt tggctatcga cttcgaaatc 1680
gaaggtgaat acctcaatt cggtacaac gacccaagag ttgacgactt ggctgttgac 1740
ttggttgaaa gatttatgaa gaagatccaa aagttgcaca cctacagaga cgctatccca 1800
acccaatctg ttttgactat cacatctaac gttgtttacg gtaagaagac tggtaacacc 1860
ccagacggta gaagagctgg tgctccattc ggtccagggt ctaacccaat gcacggtaga 1920
gacccaaaagg gtgctgtagc atctttgacc tctggtgcta agttgccatt cgcttacgct 1980
aaggacggta tctcttacac cttctctatc gttccaaacg ctttgggtaa ggacgatgaa 2040
gtagaaaga ccaacttggc tggtttgatg gacggttact tccaccacga agcatctatc 2100
gaaggtggtc aacacttgaa cgtaaatggt atgaacagag aaatgttggt ggacgctatg 2160
gaaaaccag aaaagtaccc acaattgacc atcagagttt ctggttacgc tgtagattc 2220
aactctttga ccaaggaaca acaacaagac gttatcacca gaaccttcac ccaatctatg 2280
taa 2283

```

```

<210> 3
<211> 246
<212> PRT
<213> Escherichia coli

```

```

<400> 3
Met Ser Val Ile Gly Arg Ile His Ser Phe Glu Ser Cys Gly Thr Val
1 5 10 15
Asp Gly Pro Gly Ile Arg Phe Ile Thr Phe Phe Gln Gly Cys Leu Met
20 25 30
Arg Cys Leu Tyr Cys His Asn Arg Asp Thr Trp Asp Thr His Gly Gly
35 40 45
Lys Glu Val Thr Val Glu Asp Leu Met Lys Glu Val Val Thr Tyr Arg
50 55 60
His Phe Met Asn Ala Ser Gly Gly Gly Val Thr Ala Ser Gly Gly Glu
65 70 75 80
Ala Ile Leu Gln Ala Glu Phe Val Arg Asp Trp Phe Arg Ala Cys Lys
85 90 95
Lys Glu Gly Ile His Thr Cys Leu Asp Thr Asn Gly Phe Val Arg Arg
100 105 110
Tyr Asp Pro Val Ile Asp Glu Leu Leu Glu Val Thr Asp Leu Val Met
115 120 125
Leu Asp Leu Lys Gln Met Asn Asp Glu Ile His Gln Asn Leu Val Gly
130 135 140
Val Ser Asn His Arg Thr Leu Glu Phe Ala Lys Tyr Leu Ala Asn Lys
145 150 155 160
Asn Val Lys Val Trp Ile Arg Tyr Val Val Val Pro Gly Trp Ser Asp
165 170 175
Asp Asp Asp Ser Ala His Arg Leu Gly Glu Phe Thr Arg Asp Met Gly
180 185 190

```

Asn Val Glu Lys Ile Glu Leu Leu Pro Tyr His Glu Leu Gly Lys His
 195 200 205
 Lys Trp Val Ala Met Gly Glu Glu Tyr Lys Leu Asp Gly Val Lys Pro
 210 215 220
 Pro Lys Lys Glu Thr Met Glu Arg Val Lys Gly Ile Leu Glu Gln Tyr
 225 230 235 240
 Gly His Lys Val Met Phe
 245

<210> 4
 <211> 741
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> synthetic codon optimised E.coli pflA

<400> 4
 atgtctgtta tcggtagaat ccactcttcc gaatcttgtg gtactgttga cgggtccaggt 60
 atcagattca tcaccttctt ccaagggtgt ttgatgagat gtttgtactg tcacaacaga 120
 gacacctggg acacccacgg tggtaaggaa gttactgttg aagacttgat gaaggaagtt 180
 gttacctaca gacactttat gaacgcttca ggaggtgggtg ttaccgcttc tgggtggtgaa 240
 gctatcttgc aagctgaatt tgtagagac tgggttcagag cttgtaagaa ggaaggtatc 300
 cacacctgtt tggacaccaa cggtttctgtt agaagatacg acccagttat cgacgaattg 360
 ttggaagtta ccgacttggt tatgttggac ttgaagcaaa tgaacgacga aatccaccaa 420
 aacttggttg gtgtttctaa ccacagaacc ttggaatttg ctaagtactt ggctaacaag 480
 aacgttaagg tttggatcag atacgttgtt gttccaggtt ggtctgacga cgacgactct 540
 gctcacagat tgggtgagtt caccagagac atgggtaacg ttgaaaagat cgaattgttg 600
 ccataccacg aattgggtaa gcacaagtgg gttgctatgg gtgaagaata caagttggac 660
 ggtgttaagc caccaaagaa ggaaaccatg gaaagagtta agggatatctt ggaacaatac 720
 ggtcacaagg ttatgttcta a 741

<210> 5
 <211> 376
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 5
 Met Ser Lys Gly Lys Val Leu Leu Val Leu Tyr Glu Gly Gly Lys His
 1 5 10 15
 Ala Glu Glu Gln Glu Lys Leu Leu Gly Cys Ile Glu Asn Glu Leu Gly
 20 25 30
 Ile Arg Asn Phe Ile Glu Glu Gln Gly Tyr Glu Leu Val Thr Thr Ile
 35 40 45
 Asp Lys Asp Pro Glu Pro Thr Ser Thr Val Asp Arg Glu Leu Lys Asp
 50 55 60
 Ala Glu Ile Val Ile Thr Thr Pro Phe Phe Pro Ala Tyr Ile Ser Arg
 65 70 75 80
 Asn Arg Ile Ala Glu Ala Pro Asn Leu Lys Leu Cys Val Thr Ala Gly
 85 90 95
 Val Gly Ser Asp His Val Asp Leu Glu Ala Ala Asn Glu Arg Lys Ile
 100 105 110
 Thr Val Thr Glu Val Thr Gly Ser Asn Val Val Ser Val Ala Glu His
 115 120 125

Val	Met	Ala	Thr	Ile	Leu	Val	Leu	Ile	Arg	Asn	Tyr	Asn	Gly	Gly	His
130						135					140				
Gln	Gln	Ala	Ile	Asn	Gly	Glu	Trp	Asp	Ile	Ala	Gly	Val	Ala	Lys	Asn
145					150					155					160
Glu	Tyr	Asp	Leu	Glu	Asp	Lys	Ile	Ile	Ser	Thr	Val	Gly	Ala	Gly	Arg
			165						170						175
Ile	Gly	Tyr	Arg	Val	Leu	Glu	Arg	Leu	Val	Ala	Phe	Asn	Pro	Lys	Lys
			180					185					190		
Leu	Leu	Tyr	Tyr	Asp	Tyr	Gln	Glu	Leu	Pro	Ala	Glu	Ala	Ile	Asn	Arg
		195					200					205			
Leu	Asn	Glu	Ala	Ser	Lys	Leu	Phe	Asn	Gly	Arg	Gly	Asp	Ile	Val	Gln
210						215					220				
Arg	Val	Glu	Lys	Leu	Glu	Asp	Met	Val	Ala	Gln	Ser	Asp	Val	Val	Thr
225					230					235					240
Ile	Asn	Cys	Pro	Leu	His	Lys	Asp	Ser	Arg	Gly	Leu	Phe	Asn	Lys	Lys
				245					250					255	
Leu	Ile	Ser	His	Met	Lys	Asp	Gly	Ala	Tyr	Leu	Val	Asn	Thr	Ala	Arg
			260					265					270		
Gly	Ala	Ile	Cys	Val	Ala	Glu	Asp	Val	Ala	Glu	Ala	Val	Lys	Ser	Gly
		275					280					285			
Lys	Leu	Ala	Gly	Tyr	Gly	Gly	Asp	Val	Trp	Asp	Lys	Gln	Pro	Ala	Pro
290						295					300				
Lys	Asp	His	Pro	Trp	Arg	Thr	Met	Asp	Asn	Lys	Asp	His	Val	Gly	Asn
305					310					315					320
Ala	Met	Thr	Val	His	Ile	Ser	Gly	Thr	Ser	Leu	Asp	Ala	Gln	Lys	Arg
				325					330					335	
Tyr	Ala	Gln	Gly	Val	Lys	Asn	Ile	Leu	Asn	Ser	Tyr	Phe	Ser	Lys	Lys
			340					345					350		
Phe	Asp	Tyr	Arg	Pro	Gln	Asp	Ile	Ile	Val	Gln	Asn	Gly	Ser	Tyr	Ala
		355					360					365			
Thr	Arg	Ala	Tyr	Gly	Gln	Lys	Lys								
370						375									

<210> 6
 <211> 376
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 6
 Met Ser Lys Gly Lys Val Leu Leu Val Leu Tyr Glu Gly Gly Lys His
 1 5 10 15
 Ala Glu Glu Gln Glu Lys Leu Leu Gly Cys Ile Glu Asn Glu Leu Gly
 20 25 30
 Ile Arg Asn Phe Ile Glu Glu Gln Gly Tyr Glu Leu Val Thr Thr Ile
 35 40 45
 Asp Lys Asp Pro Glu Pro Thr Ser Thr Val Asp Arg Glu Leu Lys Asp
 50 55 60
 Ala Glu Ile Val Ile Thr Thr Pro Phe Phe Pro Ala Tyr Ile Ser Arg
 65 70 75 80
 Asn Arg Ile Ala Glu Ala Pro Asn Leu Lys Leu Cys Val Thr Ala Gly
 85 90 95
 Val Gly Ser Asp His Val Asp Leu Glu Ala Ala Asn Glu Arg Lys Ile
 100 105 110
 Thr Val Thr Glu Val Thr Gly Ser Asn Val Val Ser Val Ala Glu His
 115 120 125

Val	Met	Ala	Thr	Ile	Leu	Val	Leu	Ile	Arg	Asn	Tyr	Asn	Gly	Gly	His
130						135					140				
Gln	Gln	Ala	Ile	Asn	Gly	Glu	Trp	Asp	Ile	Ala	Gly	Val	Ala	Lys	Asn
145					150					155					160
Glu	Tyr	Asp	Leu	Glu	Asp	Lys	Ile	Ile	Ser	Thr	Val	Gly	Ala	Gly	Arg
				165					170						175
Ile	Gly	Tyr	Arg	Val	Leu	Glu	Arg	Leu	Val	Ala	Phe	Asn	Pro	Lys	Lys
			180					185					190		
Leu	Leu	Tyr	Tyr	Asp	Tyr	Gln	Glu	Leu	Pro	Ala	Glu	Ala	Ile	Asn	Arg
		195					200					205			
Leu	Asn	Glu	Ala	Ser	Lys	Leu	Phe	Asn	Gly	Arg	Gly	Asp	Ile	Val	Gln
210						215					220				
Arg	Val	Glu	Lys	Leu	Glu	Asp	Met	Val	Ala	Gln	Ser	Asp	Val	Val	Thr
225					230					235					240
Ile	Asn	Cys	Pro	Leu	His	Lys	Asp	Ser	Arg	Gly	Leu	Phe	Asn	Lys	Lys
				245					250					255	
Leu	Ile	Ser	His	Met	Lys	Asp	Gly	Ala	Tyr	Leu	Val	Asn	Thr	Ala	Arg
		260					265						270		
Gly	Ala	Ile	Cys	Val	Ala	Glu	Asp	Val	Ala	Glu	Ala	Val	Lys	Ser	Gly
		275					280					285			
Lys	Leu	Ala	Gly	Tyr	Gly	Gly	Asp	Val	Trp	Asp	Lys	Gln	Pro	Ala	Pro
290						295					300				
Lys	Asp	His	Pro	Trp	Arg	Thr	Met	Asp	Asn	Lys	Asp	His	Val	Gly	Asn
305					310					315					320
Ala	Met	Thr	Val	His	Ile	Ser	Gly	Thr	Ser	Leu	Asp	Ala	Gln	Lys	Arg
				325					330					335	
Tyr	Ala	Gln	Gly	Val	Lys	Asn	Ile	Leu	Asn	Ser	Tyr	Phe	Ser	Lys	Lys
		340					345						350		
Phe	Asp	Tyr	Arg	Pro	Gln	Asp	Ile	Ile	Val	Gln	Asn	Gly	Ser	Tyr	Ala
	355					360					365				
Thr	Arg	Ala	Tyr	Gly	Gln	Lys	Lys								
370						375									

<210> 7
 <211> 316
 <212> PRT
 <213> Escherichia coli

<400> 7
 Met Ser Lys Arg Lys Val Ala Ile Ile Gly Ser Gly Asn Ile Gly Thr
 1 5 10 15
 Asp Leu Met Ile Lys Ile Leu Arg His Gly Gln His Leu Glu Met Ala
 20 25 30
 Val Met Val Gly Ile Asp Pro Gln Ser Asp Gly Leu Ala Arg Ala Arg
 35 40 45
 Arg Met Gly Val Ala Thr Thr His Glu Gly Val Ile Gly Leu Met Asn
 50 55 60
 Met Pro Glu Phe Ala Asp Ile Asp Ile Val Phe Asp Ala Thr Ser Ala
 65 70 75 80
 Gly Ala His Val Lys Asn Asp Ala Ala Leu Arg Glu Ala Lys Pro Asp
 85 90 95
 Ile Arg Leu Ile Asp Leu Thr Pro Ala Ala Ile Gly Pro Tyr Cys Val
 100 105 110
 Pro Val Val Asn Leu Glu Ala Asn Val Asp Gln Leu Asn Val Asn Met
 115 120 125

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

<210> 8
 <211> 968
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> synthetic codon optimised E.coli mhpF

<400> 8

ctgcagtcta	gatgtctaag	agaaagggtg	ctatcatcgg	ttctggtaac	atcggtactg	60
acttgatgat	caagatccta	agacacggtc	aacacttgga	aatggctggt	atggttggta	120
tcgacccaca	atctgacggg	ttggctagag	ctagaagaat	gggtggtgct	accacccacg	180
aaggtggtat	cggtttgatg	aacatgccag	aattcgctga	catcgacatc	gttttcgacg	240
ctacctctgc	tggtgctcac	gttaagaacg	acgctgcttt	gagagaagct	aagccagaca	300
tcagattgat	cgacttgacc	ccagctgcta	tcggtccata	ctgtgttcca	gttggttaact	360
tggaagctaa	cgttgaccaa	ttaaacgtta	acatgggttac	ctgtgggtgg	caagctacca	420
tcccaatggg	tgctgctggt	tcaagagttg	ctagagttca	ctacgctgaa	atcatcgctt	480
ctatcgcttc	taagtctgct	gggccaggta	ccagagctaa	catcgacgaa	ttcaccgaaa	540
ccacctctag	ggctatcgaa	gttggttggtg	gtgctgctaa	gggtaaggct	atcatcgctt	600
tgaacccagc	tgaaccacca	ttgatgatga	gagacaccgt	ttacgttttg	tctgacgaag	660
catctcaaga	cgacatcgaa	gcttcaatca	acgaaatggc	tgaagctggt	caagcatacg	720
ttccagggtta	cagattgaag	caaagagttc	aattcgaagt	tatcccacaa	gacaagccag	780
ttactttgcc	aggtgttggt	caattctctg	gtttgaagac	cgctgttttg	ttggaagttg	840
aaggtgctgc	tactacttg	ccagcttacg	ctggtaactt	ggacattatg	acctcttctg	900
ctttggctac	cgctgaaaag	atggctcaat	ctttggctag	aaaggctggt	gaagctgctt	960
aagcgcgc						968

<210> 9
 <211> 891

<212> PRT

<213> Escherichia coli

<400> 9

Met	Ala	Val	Thr	Asn	Val	Ala	Glu	Leu	Asn	Ala	Leu	Val	Glu	Arg	Val
1				5					10					15	
Lys	Lys	Ala	Gln	Arg	Glu	Tyr	Ala	Ser	Phe	Thr	Gln	Glu	Gln	Val	Asp
			20					25					30		
Lys	Ile	Phe	Arg	Ala	Ala	Ala	Leu	Ala	Ala	Ala	Asp	Ala	Arg	Ile	Pro
		35					40					45			
Leu	Ala	Lys	Met	Ala	Val	Ala	Glu	Ser	Gly	Met	Gly	Ile	Val	Glu	Asp
	50					55				60					
Lys	Val	Ile	Lys	Asn	His	Phe	Ala	Ser	Glu	Tyr	Ile	Tyr	Asn	Ala	Tyr
65				70					75					80	
Lys	Asp	Glu	Lys	Thr	Cys	Gly	Val	Leu	Ser	Glu	Asp	Asp	Thr	Phe	Gly
				85					90					95	
Thr	Ile	Thr	Ile	Ala	Glu	Pro	Ile	Gly	Ile	Ile	Cys	Gly	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	Ile	Phe	Ser	Pro	His	Pro	Arg	Ala	Lys	Asp
	130					135					140				
Ala	Thr	Asn	Lys	Ala	Ala	Asp	Ile	Val	Leu	Gln	Ala	Ala	Ile	Ala	Ala
145					150					155					160
Gly	Ala	Pro	Lys	Asp	Leu	Ile	Gly	Trp	Ile	Asp	Gln	Pro	Ser	Val	Glu
				165					170					175	
Leu	Ser	Asn	Ala	Leu	Met	His	His	Pro	Asp	Ile	Asn	Leu	Ile	Leu	Ala
		180						185				190			
Thr	Gly	Gly	Pro	Gly	Met	Val	Lys	Ala	Ala	Tyr	Ser	Ser	Gly	Lys	Pro
		195					200					205			
Ala	Ile	Gly	Val	Gly	Ala	Gly	Asn	Thr	Pro	Val	Val	Ile	Asp	Glu	Thr
	210					215					220				
Ala	Asp	Ile	Lys	Arg	Ala	Val	Ala	Ser	Val	Leu	Met	Ser	Lys	Thr	Phe
225					230					235					240
Asp	Asn	Gly	Val	Ile	Cys	Ala	Ser	Glu	Gln	Ser	Val	Val	Val	Val	Asp
				245					250					255	
Ser	Val	Tyr	Asp	Ala	Val	Arg	Glu	Arg	Phe	Ala	Thr	His	Gly	Gly	Tyr
		260						265					270		
Leu	Leu	Gln	Gly	Lys	Glu	Leu	Lys	Ala	Val	Gln	Asp	Val	Ile	Leu	Lys
		275					280					285			
Asn	Gly	Ala	Leu	Asn	Ala	Ala	Ile	Val	Gly	Gln	Pro	Ala	Tyr	Lys	Ile
	290					295					300				
Ala	Glu	Leu	Ala	Gly	Phe	Ser	Val	Pro	Glu	Asn	Thr	Lys	Ile	Leu	Ile
305					310					315					320
Gly	Glu	Val	Thr	Val	Val	Asp	Glu	Ser	Glu	Pro	Phe	Ala	His	Glu	Lys
				325					330					335	
Leu	Ser	Pro	Thr	Leu	Ala	Met	Tyr	Arg	Ala	Lys	Asp	Phe	Glu	Asp	Ala
			340					345					350		
Val	Glu	Lys	Ala	Glu	Lys	Leu	Val	Ala	Met	Gly	Gly	Ile	Gly	His	Thr
		355				360						365			
Ser	Cys	Leu	Tyr	Thr	Asp	Gln	Asp	Asn	Gln	Pro	Ala	Arg	Val	Ser	Tyr
	370					375					380				
Phe	Gly	Gln	Lys	Met	Lys	Thr	Ala	Arg	Ile	Leu	Ile	Asn	Thr	Pro	Ala
385					390					395					400
Ser	Gln	Gly	Gly	Ile	Gly	Asp	Leu	Tyr	Asn	Phe	Lys	Leu	Ala	Pro	Ser
				405					410					415	

Leu	Thr	Leu	Gly	Cys	Gly	Ser	Trp	Gly	Gly	Asn	Ser	Ile	Ser	Glu	Asn
			420					425					430		
Val	Gly	Pro	Lys	His	Leu	Ile	Asn	Lys	Lys	Thr	Val	Ala	Lys	Arg	Ala
		435					440					445			
Glu	Asn	Met	Leu	Trp	His	Lys	Leu	Pro	Lys	Ser	Ile	Tyr	Phe	Arg	Arg
	450					455					460				
Gly	Ser	Leu	Pro	Ile	Ala	Leu	Asp	Glu	Val	Ile	Thr	Asp	Gly	His	Lys
465					470					475					480
Arg	Ala	Leu	Ile	Val	Thr	Asp	Arg	Phe	Leu	Phe	Asn	Asn	Gly	Tyr	Ala
				485					490					495	
Asp	Gln	Ile	Thr	Ser	Val	Leu	Lys	Ala	Gly	Val	Glu	Thr	Glu	Val	
			500					505				510			
Phe	Phe	Glu	Val	Glu	Ala	Asp	Pro	Thr	Leu	Ser	Ile	Val	Arg	Lys	Gly
		515					520					525			
Ala	Glu	Leu	Ala	Asn	Ser	Phe	Lys	Pro	Asp	Val	Ile	Ile	Ala	Leu	Gly
	530					535					540				
Gly	Gly	Ser	Pro	Met	Asp	Ala	Ala	Lys	Ile	Met	Trp	Val	Met	Tyr	Glu
545					550					555					560
His	Pro	Glu	Thr	His	Phe	Glu	Glu	Leu	Ala	Leu	Arg	Phe	Met	Asp	Ile
				565					570					575	
Arg	Lys	Arg	Ile	Tyr	Lys	Phe	Pro	Lys	Met	Gly	Val	Lys	Ala	Lys	Met
			580					585					590		
Ile	Ala	Val	Thr	Thr	Thr	Ser	Gly	Thr	Gly	Ser	Glu	Val	Thr	Pro	Phe
		595					600					605			
Ala	Val	Val	Thr	Asp	Asp	Ala	Thr	Gly	Gln	Lys	Tyr	Pro	Leu	Ala	Asp
	610					615					620				
Tyr	Ala	Leu	Thr	Pro	Asp	Met	Ala	Ile	Val	Asp	Ala	Asn	Leu	Val	Met
625					630					635					640
Asp	Met	Pro	Lys	Ser	Leu	Cys	Ala	Phe	Gly	Gly	Leu	Asp	Ala	Val	Thr
				645					650					655	
His	Ala	Met	Glu	Ala	Tyr	Val	Ser	Val	Leu	Ala	Ser	Glu	Phe	Ser	Asp
		660						665					670		
Gly	Gln	Ala	Leu	Gln	Ala	Leu	Lys	Leu	Leu	Lys	Glu	Tyr	Leu	Pro	Ala
		675					680					685			
Ser	Tyr	His	Glu	Gly	Ser	Lys	Asn	Pro	Val	Ala	Arg	Glu	Arg	Val	His
	690					695					700				
Ser	Ala	Ala	Thr	Ile	Ala	Gly	Ile	Ala	Phe	Ala	Asn	Ala	Phe	Leu	Gly
705					710					715					720
Val	Cys	His	Ser	Met	Ala	His	Lys	Leu	Gly	Ser	Gln	Phe	His	Ile	Pro
				725					730					735	
His	Gly	Leu	Ala	Asn	Ala	Leu	Leu	Ile	Cys	Asn	Val	Ile	Arg	Tyr	Asn
		740						745					750		
Ala	Asn	Asp	Asn	Pro	Thr	Lys	Gln	Thr	Ala	Phe	Ser	Gln	Tyr	Asp	Arg
		755					760					765			
Pro	Gln	Ala	Arg	Arg	Arg	Tyr	Ala	Glu	Ile	Ala	Asp	His	Leu	Gly	Leu
	770					775						780			
Ser	Ala	Pro	Gly	Asp	Arg	Thr	Ala	Ala	Lys	Ile	Glu	Lys	Leu	Leu	Ala
785					790					795					800
Trp	Leu	Glu	Thr	Leu	Lys	Ala	Glu	Leu	Gly	Ile	Pro	Lys	Ser	Ile	Arg
				805					810					815	
Glu	Ala	Gly	Val	Gln	Glu	Ala	Asp	Phe	Leu	Ala	Asn	Val	Asp	Lys	Leu
			820					825					830		
Ser	Glu	Asp	Ala	Phe	Asp	Asp	Gln	Cys	Thr	Gly	Ala	Asn	Pro	Arg	Tyr
		835					840					845			
Pro	Leu	Ile	Ser	Glu	Leu	Lys	Gln	Ile	Leu	Leu	Asp	Thr	Tyr	Tyr	Gly
	850					855					860				

Arg Asp Tyr Val Glu Gly Glu Thr Ala Ala Lys Lys Glu Ala Ala Pro
 865 870 875 880
 Ala Lys Ala Glu Lys Lys Ala Lys Lys Ser Ala
 885 890

<210> 10
 <211> 2690
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> synthetic codon optimised E.coli adhE

<400> 10
 tctagaaaat ggctggtacc aacggtgctg aattgaacgc tttggttgaa agggttaaga 60
 aggc tcaaag agaatacgt tctttcaccc aagaacaagt tgacaagatc ttcagagctg 120
 ctgctttggc tgctgctgac gctagaatcc cattggctaa gatggctggt gctgaatctg 180
 gtatgggtat cgttgaagac aaggttatca agaaccactt cgcttctgaa tacatctaca 240
 acgcttacaa ggacgaaaag acctgtggtg ttttgtcaga agacgacacc ttcggtacca 300
 tcaccatcgc tgaaccaatc ggtatcatct gtggtatcgt tccaaccacc aaccaaacct 360
 ctaccgctat cttcaagtct ttgatctctt tgaagaccag aaacgctatc atcttctctc 420
 cacaccaag agctaaagac gctaccaaca aggc tgc tga categ ttttg caagctgcta 480
 tcgctgctgg tgctccaaag gacttgatcg gttggatcga ccaaccatct gttgaattgt 540
 ctaacgcttt gatgcaccac ccagacatca acttgatctt ggctaccggt ggtccaggta 600
 tggttaaggc tgcttactct tctggtaagc cagctatcgg tgttggtgct ggtaacaccc 660
 cagttgttat cgacgaaacc gctgacatca agagagctgt tgcttctggt ttgatgtcta 720
 agaccttcga caacggtggt atctgtgctt ctgaacaatc tgttggtggt gttgactctg 780
 tttacgacgc tgtagagaa agatteccta cccacggtgg ttacttggtg caaggtaagg 840
 aattgaaggc tgttcaagac gttatcttga agaacggtgc tttgaacgct gctatcggtg 900
 gtcaaccagc ttacaagatc gctgaattag ctggtttctc tgttccagaa aacaccaaga 960
 tcttgatcgg tgaagttacc gttgttgacg aatctgaacc attcgctcac gaaaagttgt 1020
 ctccaacctt ggctatgtac agagctaagg acttcgaaga cgctgttgaa aaagctgaaa 1080
 agttgggttg tatgggtggt attggtcaca cctcttggtt gtacaccgac caagacaacc 1140
 aaccagctag agtttcttac ttcggtcaaa agatgaagac cgctagaatc ttgatcaaca 1200
 cccagcttc tcaaggtggt atcgggtgact tgtacaactt caagttggct ccatctttga 1260
 ccttgggttg tggttcttgg ggtggtaact ctatctctga aaacggtggt ccaaagcact 1320
 tgatcaacaa gaagaccgtt gctaagagag ctgaaaacat gttgtggcac aagttgcaa 1380
 aatctatcta cttcagaaga ggttctttgc caatcgcttt ggacgaagtt atcaccgacg 1440
 gtcacaagag agctttgatc gttaccgaca gattcttggt caacaacggt tacgctgacc 1500
 aaatcacctc tgttttgaag gctgctggtg ttgaaaccga agttttcttc gaagttgaag 1560
 ctgacccaac cttgtctatc gttagaaagg gtgctgaatt ggctaactct ttcaagccag 1620
 acgttatcat cgctttgggt ggtggttctc caatggacgc tgctaagatc atgtgggtta 1680
 tgtacgaaca cccagaaacc cacttcgaag aattggcttt gagattcatg gacatcagaa 1740
 agagaatcta caagttccca aagatgggtg ttaaggctaa gatgatcgct gttaccacca 1800
 cctctggtac cggttctgaa gttaccccat tcgctggtgt taccgacgac gctaccggtc 1860
 aaaagtaccc attggctgac tacgctttga cccagacat ggctatcggt gacgctaact 1920
 tggttatgga catgccaaag tctttgtgtg ctttcggtgg tttggacgct gttaccacag 1980
 ctatggaagc ttacgtttct gttttggctt ctgaattctc tgacgggtcaa gctttgcaag 2040
 ctttgaagtt gttgaaggaa tacttgccag cttcttacca cgaaggttct aagaaccag 2100
 ttgctagaga aagagttcac tctgctgcta ccacgctgg tatoctttc gctaacgctt 2160
 tcttgggtgt ttgtcactct atggctcaca agttgggttc tcaattccac atcccacag 2220
 gtttggctaa cgctttgttg atctgtaacg ttatcagata caacgctaac gacaacccaa 2280
 ccaagcaaac cgctttctct caatacgaca gaccacaagc tagaagaaga tacgctgaaa 2340
 tcgctgacca cttgggtttg tctgctccag gtgacagaac cgctgcaaag atcgaaaagt 2400
 tgttggcttg gttgaaacc ttgaaggctg aattgggtat cccaaagtct atcagagaag 2460

```

ctggtgttca agaagctgac ttcttggtta acgttgacaa gttgtctgaa gacgctttcg 2520
acgaccaatg taccggtgct aacccaagat acccattgat ctctgaattg aagcaaattc 2580
tggttgacac ctactacggt agagactacg ttgaaggtga aacogctgct aagaaggaag 2640
ctgctccagc taaggctgaa aagaaggcta agaagtctgc ttagcttaag 2690

```

```

<210> 11
<211> 870
<212> PRT
<213> Entamoeba histolytica

```

```

<400> 11
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		
385					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		
465					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		
545					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		
625					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		
705					710					715					720		
Phe	Ala	Ser	Ala	Phe	Leu	Gly	Met	Asp	His	Ser	Met	Ala	His	Lys	Val		
				725					730					735			
Gly	Ala	Ala	Phe	His	Leu	Pro	His	Gly	Arg	Cys	Val	Ala	Val	Leu	Leu		
			740					745					750				
Pro	His	Val	Ile	Arg	Tyr	Asn	Gly	Gln	Lys	Pro	Arg	Lys	Leu	Ala	Met		
		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					795					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> 12

<211> 2627

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic codon optimised E.histolytica ADH2

<400> 12

tctagaaaat	gtctacccaa	caaaccatga	ccgttgacga	acacatcaac	caattagtta	60
gaaaggctca	agttgctttg	aaggaatact	tgaagccaga	atacacccaa	gaaaagatcg	120
actacatcgt	taagaaggct	tctgttgctg	ctttggacca	acactgtgct	ttggctgctg	180
ctgctgttga	agaaaccggt	agaggtatct	tcgaagacaa	ggctaccaag	aacatcttcg	240
cttgtgaaca	cggtacccac	gaaatgagac	acgctaagac	cgttggtatc	atcaacgttg	300
acccattgta	cggtatcacc	gaaatcgctg	aaccagttgg	tgttgtttgt	ggtggttacc	360
cagttacca	cccaacctct	accgctatct	tcaagtcttt	gatctctatc	aagaccagaa	420
acccaatcgt	tttctctttc	cacctatctg	ctttgaagtg	ttctattatg	gctgctaaaa	480
tcggttagaga	cgctgctatc	gctgctgggt	ctccagaaaa	ctgtatccaa	tggatcgaat	540
tcggtgggtat	cgaagcttct	aacaagttga	tgaaccaccc	aggtggtgct	actatcttgg	600
ctaccggtgg	taacgctatg	gttaaggctg	catactcttc	tggtaagcca	gctttgggtg	660
ttggtgctgg	taacgttcca	acctacatcg	aaaagacctg	taacatcaag	caagctgcta	720
acgacgttgt	tatgtctaag	tcttttcgaca	acggtatgat	ctgtgcttct	gaacaagctg	780
ctatcatcga	caaggaaatc	tacgaccaag	ttgttgaaga	aatgaagacc	ttgggtgctt	840
acttcatcaa	cgaagaagaa	aaggctaagt	tggaaaagtt	catgttcggt	gttaacgctt	900
actctgctga	cgtaacaac	gctagattga	acccaaagtg	tccaggtatg	tctccacaat	960
ggttcgctga	acaagttgg	atcaaggtag	cagaagactg	taacatcatc	tgtgctgttt	1020
gtaaggaagt	tgggtccaaac	gaaccattga	ccagagaaaa	gttgtctcca	gttttggtta	1080
tcttgaaagc	tgaaaacacc	caagacggta	tcgacaaggc	tgaagctatg	gttgaattta	1140
acggtagagg	tcactctgct	gctatccact	ctaacgacaa	ggctggtggt	gaaaagtacg	1200
ctttgaccat	gaaggcttgt	agaatcttgc	acaacacccc	atcttctcaa	ggtggtatcg	1260
gttctatcta	caactacatc	tggccatctt	tcaccttggg	ttgtggttct	tacggtggta	1320
actctgtttc	tgttaacggt	acctaccaca	acttggtgaa	catcaagaga	ttggctgaca	1380
gaagaaacaa	cttgcaatgg	ttcagagttc	caccaaagat	cttcttcgaa	ccacactota	1440
tcagatactt	ggctgaattg	aaggaattgt	ctaagatctt	catcgtttct	gacagaatga	1500
tgtacaagtt	gggttacggt	gacagagtta	tggacgtttt	gaagagaaga	tctaacgaag	1560
ttgaaatcga	aatcttcatc	gacgttgaac	cagacccatc	tatccaaacc	gttcaaaagg	1620
gtttggctgt	tatgaacacc	ttcgggtccag	acaacatcat	cgctatcggt	ggtggttctg	1680
ctatggacgc	tgttaagatc	atgtggttgt	tgtacgaaca	cccagaagct	gacttcttcg	1740
ctatgaagca	aaagttcatc	gacttgagaa	agagagcttt	caagttccca	accatgggta	1800
agaaggctag	attgatctgt	atcccaacca	cctctggtac	cggttctgaa	gttaccocat	1860
tcgctgttat	ctctgaccac	gaaaceggta	agaagtaccc	attggctgac	tactctttga	1920
ccccatctgt	tgtatcgtt	gacccaatgt	tcaccatgtc	tttgccaaag	agagctatcg	1980
ctgacaccgg	tttggtacgtt	ttggttcacg	ctaccgaagc	ttacgtttct	gttatggcta	2040
acgaatacac	cgacggtttg	gctagagaag	ctgttaagtt	ggtttttgaa	aacttggtga	2100

```

agtcttataa cggtagcttg gaagctagag aaaagatgca caacgctgct accatcgctg 2160
gtatggcttt cgcttctgct ttcttgggta tggaccactc tatggctcac aagggttggtg 2220
ctgctttcca cttgccacac ggtagatgtg ttgctgtttt gttgccacac gttatcagat 2280
acaacgggtca aaagccaaga aagttggcta tgtggccaaa gtacaacttc tacaaggctg 2340
accaaagata catggaattg gctcaaattg ttggtttgaa gtgtaacacc ccagctgaag 2400
gtgttgaagc ttctgctaag gcttgtgaag aattgatgaa ggctaccgaa accatcaccg 2460
gtttcaagaa ggctaacatc gacgaagctg cttggatgtc taaggttcca gaaatggctt 2520
tggttgcttt cgaagaccaa tggtctccag ctaaccaag agttccaatg gtttaaggaca 2580
tggaagagat cttgaaggct gcttactacc caatcgctta gcttaag 2627

```

```

<210> 13
<211> 312
<212> PRT
<213> Saccharomyces cerevisiae

```

```

<400> 13
Met Pro Ala Thr Leu His Asp Ser Thr Lys Ile Leu Ser Leu Asn Thr
1          5          10          15
Gly Ala Gln Ile Pro Gln Ile Gly Leu Gly Thr Trp Gln Ser Lys Glu
20          25          30
Asn Asp Ala Tyr Lys Ala Val Leu Thr Ala Leu Lys Asp Gly Tyr Arg
35          40          45
His Ile Asp Thr Ala Ala Ile Tyr Arg Asn Glu Asp Gln Val Gly Gln
50          55          60
Ala Ile Lys Asp Ser Gly Val Pro Arg Glu Glu Ile Phe Val Thr Thr
65          70          75          80
Lys Leu Trp Cys Thr Gln His His Glu Pro Glu Val Ala Leu Asp Gln
85          90          95
Ser Leu Lys Arg Leu Gly Leu Asp Tyr Val Asp Leu Tyr Leu Met His
100         105         110
Trp Pro Ala Arg Leu Asp Pro Ala Tyr Ile Lys Asn Glu Asp Ile Leu
115         120         125
Ser Val Pro Thr Lys Lys Asp Gly Ser Arg Ala Val Asp Ile Thr Asn
130         135         140
Trp Asn Phe Ile Lys Thr Trp Glu Leu Met Gln Glu Leu Pro Lys Thr
145         150         155         160
Gly Lys Thr Lys Ala Val Gly Val Ser Asn Phe Ser Ile Asn Asn Leu
165         170         175
Lys Asp Leu Leu Ala Ser Gln Gly Asn Lys Leu Thr Pro Ala Ala Asn
180         185         190
Gln Val Glu Ile His Pro Leu Leu Pro Gln Asp Glu Leu Ile Asn Phe
195         200         205
Cys Lys Ser Lys Gly Ile Val Val Glu Ala Tyr Ser Pro Leu Gly Ser
210         215         220
Thr Asp Ala Pro Leu Leu Lys Glu Pro Val Ile Leu Glu Ile Ala Lys
225         230         235         240
Lys Asn Asn Val Gln Pro Gly His Val Val Ile Ser Trp His Val Gln
245         250         255
Arg Gly Tyr Val Val Leu Pro Lys Ser Val Asn Pro Asp Arg Ile Lys
260         265         270
Thr Asn Arg Lys Ile Phe Thr Leu Ser Thr Glu Asp Phe Glu Ala Ile
275         280         285

```

Asn Asn Ile Ser Lys Glu Lys Gly Glu Lys Arg Val Val His Pro Asn
 290 295 300
 Trp Ser Pro Phe Glu Val Phe Lys
 305 310

<210> 14
 <211> 584
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 14
 Met Ser Ala Lys Ser Phe Glu Val Thr Asp Pro Val Asn Ser Ser Leu
 1 5 10 15
 Lys Gly Phe Ala Leu Ala Asn Pro Ser Ile Thr Leu Val Pro Glu Glu
 20 25 30
 Lys Ile Leu Phe Arg Lys Thr Asp Ser Asp Lys Ile Ala Leu Ile Ser
 35 40 45
 Gly Gly Gly Ser Gly His Glu Pro Thr His Ala Gly Phe Ile Gly Lys
 50 55 60
 Gly Met Leu Ser Gly Ala Val Val Gly Glu Ile Phe Ala Ser Pro Ser
 65 70 75 80
 Thr Lys Gln Ile Leu Asn Ala Ile Arg Leu Val Asn Glu Asn Ala Ser
 85 90 95
 Gly Val Leu Leu Ile Val Lys Asn Tyr Thr Gly Asp Val Leu His Phe
 100 105 110
 Gly Leu Ser Ala Glu Arg Ala Arg Ala Leu Gly Ile Asn Cys Arg Val
 115 120 125
 Ala Val Ile Gly Asp Asp Val Ala Val Gly Arg Glu Lys Gly Gly Met
 130 135 140
 Val Gly Arg Arg Ala Leu Ala Gly Thr Val Leu Val His Lys Ile Val
 145 150 155 160
 Gly Ala Phe Ala Glu Glu Tyr Ser Ser Lys Tyr Gly Leu Asp Gly Thr
 165 170 175
 Ala Lys Val Ala Lys Ile Ile Asn Asp Asn Leu Val Thr Ile Gly Ser
 180 185 190
 Ser Leu Asp His Cys Lys Val Pro Gly Arg Lys Phe Glu Ser Glu Leu
 195 200 205
 Asn Glu Lys Gln Met Glu Leu Gly Met Gly Ile His Asn Glu Pro Gly
 210 215 220
 Val Lys Val Leu Asp Pro Ile Pro Ser Thr Glu Asp Leu Ile Ser Lys
 225 230 235 240
 Tyr Met Leu Pro Lys Leu Leu Asp Pro Asn Asp Lys Asp Arg Ala Phe
 245 250 255
 Val Lys Phe Asp Glu Asp Asp Glu Val Val Leu Leu Val Asn Asn Leu
 260 265 270
 Gly Gly Val Ser Asn Phe Val Ile Ser Ser Ile Thr Ser Lys Thr Thr
 275 280 285
 Asp Phe Leu Lys Glu Asn Tyr Asn Ile Thr Pro Val Gln Thr Ile Ala
 290 295 300
 Gly Thr Leu Met Thr Ser Phe Asn Gly Asn Gly Phe Ser Ile Thr Leu
 305 310 315 320
 Leu Asn Ala Thr Lys Ala Thr Lys Ala Leu Gln Ser Asp Phe Glu Glu
 325 330 335
 Ile Lys Ser Val Leu Asp Leu Leu Asn Ala Phe Thr Asn Ala Pro Gly
 340 345 350

Trp Pro Ile Ala Asp Phe Glu Lys Thr Ser Ala Pro Ser Val Asn Asp
 355 360 365
 Asp Leu Leu His Asn Glu Val Thr Ala Lys Ala Val Gly Thr Tyr Asp
 370 375 380
 Phe Asp Lys Phe Ala Glu Trp Met Lys Ser Gly Ala Glu Gln Val Ile
 385 390 395 400
 Lys Ser Glu Pro His Ile Thr Glu Leu Asp Asn Gln Val Gly Asp Gly
 405 410 415
 Asp Cys Gly Tyr Thr Leu Val Ala Gly Val Lys Gly Ile Thr Glu Asn
 420 425 430
 Leu Asp Lys Leu Ser Lys Asp Ser Leu Ser Gln Ala Val Ala Gln Ile
 435 440 445
 Ser Asp Phe Ile Glu Gly Ser Met Gly Gly Thr Ser Gly Gly Leu Tyr
 450 455 460
 Ser Ile Leu Leu Ser Gly Phe Ser His Gly Leu Ile Gln Val Cys Lys
 465 470 475 480
 Ser Lys Asp Glu Pro Val Thr Lys Glu Ile Val Ala Lys Ser Leu Gly
 485 490 495
 Ile Ala Leu Asp Thr Leu Tyr Lys Tyr Thr Lys Ala Arg Lys Gly Ser
 500 505 510
 Ser Thr Met Ile Asp Ala Leu Glu Pro Phe Val Lys Glu Phe Thr Ala
 515 520 525
 Ser Lys Asp Phe Asn Lys Ala Val Lys Ala Ala Glu Glu Gly Ala Lys
 530 535 540
 Ser Thr Ala Thr Phe Glu Ala Lys Phe Gly Arg Ala Ser Tyr Val Gly
 545 550 555 560
 Asp Ser Ser Gln Val Glu Asp Pro Gly Ala Val Gly Leu Cys Glu Phe
 565 570 575
 Leu Lys Gly Val Gln Ser Ala Leu
 580

<210> 15
 <211> 591
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 15
 Met Ser His Lys Gln Phe Lys Ser Asp Gly Asn Ile Val Thr Pro Tyr
 1 5 10 15
 Leu Leu Gly Leu Ala Arg Ser Asn Pro Gly Leu Thr Val Ile Lys His
 20 25 30
 Asp Arg Val Val Phe Arg Thr Ala Ser Ala Pro Asn Ser Gly Asn Pro
 35 40 45
 Pro Lys Val Ser Leu Val Ser Gly Gly Gly Ser Gly His Glu Pro Thr
 50 55 60
 His Ala Gly Phe Val Gly Glu Gly Ala Leu Asp Ala Ile Ala Ala Gly
 65 70 75 80
 Ala Ile Phe Ala Ser Pro Ser Thr Lys Gln Ile Tyr Ser Ala Ile Lys
 85 90 95
 Ala Val Glu Ser Pro Lys Gly Thr Leu Ile Ile Val Lys Asn Tyr Thr
 100 105 110
 Gly Asp Ile Ile His Phe Gly Leu Ala Ala Glu Arg Ala Lys Ala Ala
 115 120 125
 Gly Met Lys Val Glu Leu Val Ala Val Gly Asp Asp Val Ser Val Gly
 130 135 140

Lys	Lys	Lys	Gly	Ser	Leu	Val	Gly	Arg	Arg	Gly	Leu	Gly	Ala	Thr	Val
145					150					155					160
Leu	Val	His	Lys	Ile	Ala	Gly	Ala	Ala	Ala	Ser	His	Gly	Leu	Glu	Leu
				165						170					175
Ala	Glu	Val	Ala	Glu	Val	Ala	Gln	Ser	Val	Val	Asp	Asn	Ser	Val	Thr
				180						185					190
Ile	Ala	Ala	Ser	Leu	Asp	His	Cys	Thr	Val	Pro	Gly	His	Lys	Pro	Glu
			195					200							205
Ala	Ile	Leu	Gly	Glu	Asn	Glu	Tyr	Glu	Ile	Gly	Met	Gly	Ile	His	Asn
	210						215					220			
Glu	Ser	Gly	Thr	Tyr	Lys	Ser	Ser	Pro	Leu	Pro	Ser	Ile	Ser	Glu	Leu
225					230					235					240
Val	Ser	Gln	Met	Leu	Pro	Leu	Leu	Leu	Asp	Glu	Asp	Glu	Asp	Arg	Ser
				245						250					255
Tyr	Val	Lys	Phe	Glu	Pro	Lys	Glu	Asp	Val	Val	Leu	Met	Val	Asn	Asn
			260					265							270
Met	Gly	Gly	Met	Ser	Asn	Leu	Glu	Leu	Gly	Tyr	Ala	Ala	Glu	Val	Ile
			275					280							285
Ser	Glu	Gln	Leu	Ile	Asp	Lys	Tyr	Gln	Ile	Val	Pro	Lys	Arg	Thr	Ile
	290						295								300
Thr	Gly	Ala	Phe	Ile	Thr	Ala	Leu	Asn	Gly	Pro	Gly	Phe	Gly	Ile	Thr
305					310					315					320
Leu	Met	Asn	Ala	Ser	Lys	Ala	Gly	Gly	Asp	Ile	Leu	Lys	Tyr	Phe	Asp
				325						330					335
Tyr	Pro	Thr	Thr	Ala	Ser	Gly	Trp	Asn	Gln	Met	Tyr	His	Ser	Ala	Lys
			340						345						350
Asp	Trp	Glu	Val	Leu	Ala	Lys	Gly	Gln	Val	Pro	Thr	Ala	Pro	Ser	Leu
			355					360							365
Lys	Thr	Leu	Arg	Asn	Glu	Lys	Gly	Ser	Gly	Val	Lys	Ala	Asp	Tyr	Asp
	370						375								380
Thr	Phe	Ala	Lys	Ile	Leu	Leu	Ala	Gly	Ile	Ala	Lys	Ile	Asn	Glu	Val
385					390					395					400
Glu	Pro	Lys	Val	Thr	Trp	Tyr	Asp	Thr	Ile	Ala	Gly	Asp	Gly	Asp	Cys
				405						410					415
Gly	Thr	Thr	Leu	Val	Ser	Gly	Gly	Glu	Ala	Leu	Glu	Glu	Ala	Ile	Lys
			420						425						430
Asn	His	Thr	Leu	Arg	Leu	Glu	Asp	Ala	Ala	Leu	Gly	Ile	Glu	Asp	Ile
			435					440							445
Ala	Tyr	Met	Val	Glu	Asp	Ser	Met	Gly	Gly	Thr	Ser	Gly	Gly	Leu	Tyr
	450						455								460
Ser	Ile	Tyr	Leu	Ser	Ala	Leu	Ala	Gln	Gly	Val	Arg	Asp	Ser	Gly	Asp
465					470					475					480
Lys	Glu	Leu	Thr	Ala	Glu	Thr	Phe	Lys	Lys	Ala	Ser	Asn	Val	Ala	Leu
				485						490					495
Asp	Ala	Leu	Tyr	Lys	Tyr	Thr	Arg	Ala	Arg	Pro	Gly	Tyr	Arg	Thr	Leu
			500						505						510
Ile	Asp	Ala	Leu	Gln	Pro	Phe	Val	Glu	Ala	Leu	Lys	Ala	Gly	Lys	Gly
			515						520						525
Pro	Arg	Ala	Ala	Ala	Gln	Ala	Ala	Tyr	Asp	Gly	Ala	Glu	Lys	Thr	Arg
	530								535						540
Lys	Met	Asp	Ala	Leu	Val	Gly	Arg	Ala	Ser	Tyr	Val	Ala	Lys	Glu	Glu
545					550					555					560
Leu	Arg	Lys	Leu	Asp	Ser	Glu	Gly	Gly	Leu	Pro	Asp	Pro	Gly	Ala	Val
				565						570					575
Gly	Leu	Ala	Ala	Leu	Leu	Asp	Gly	Phe	Val	Thr	Ala	Ala	Gly	Tyr	
			580						585						590

<210> 16
 <211> 560
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 16

Met	Ser	Leu	Ile	Ser	Ile	Leu	Ser	Pro	Leu	Ile	Thr	Ser	Glu	Gly	Leu
1				5					10					15	
Asp	Ser	Arg	Ile	Lys	Pro	Ser	Pro	Lys	Lys	Asp	Ala	Ser	Thr	Thr	Thr
			20					25					30		
Lys	Pro	Ser	Leu	Trp	Lys	Thr	Thr	Glu	Phe	Lys	Phe	Tyr	Tyr	Ile	Ala
		35					40					45			
Phe	Leu	Val	Val	Val	Pro	Leu	Met	Phe	Tyr	Ala	Gly	Leu	Gln	Ala	Ser
	50					55					60				
Ser	Pro	Glu	Asn	Pro	Asn	Tyr	Ala	Arg	Tyr	Glu	Arg	Leu	Leu	Ser	Gln
65					70					75					80
Gly	Trp	Leu	Phe	Gly	Arg	Lys	Val	Asp	Asn	Ser	Asp	Ser	Gln	Tyr	Arg
				85				90					95		
Phe	Phe	Arg	Asp	Asn	Phe	Ala	Leu	Leu	Ser	Val	Leu	Met	Leu	Val	His
			100					105					110		
Thr	Ser	Ile	Lys	Arg	Ile	Val	Leu	Tyr	Ser	Thr	Asn	Ile	Thr	Lys	Leu
		115					120					125			
Arg	Phe	Asp	Leu	Ile	Phe	Gly	Leu	Ile	Phe	Leu	Val	Ala	Ala	His	Gly
	130					135					140				
Val	Asn	Ser	Ile	Arg	Ile	Leu	Ala	His	Met	Leu	Ile	Leu	Tyr	Ala	Ile
145					150					155					160
Ala	His	Val	Leu	Lys	Asn	Phe	Arg	Arg	Ile	Ala	Thr	Ile	Ser	Ile	Trp
				165				170					175		
Ile	Tyr	Gly	Ile	Ser	Thr	Leu	Phe	Ile	Asn	Asp	Asn	Phe	Arg	Ala	Tyr
		180					185					190			
Pro	Phe	Gly	Asn	Ile	Cys	Ser	Phe	Leu	Ser	Pro	Leu	Asp	His	Trp	Tyr
	195					200					205				
Arg	Gly	Ile	Ile	Pro	Arg	Trp	Asp	Val	Phe	Phe	Asn	Phe	Thr	Leu	Leu
	210					215					220				
Arg	Val	Leu	Ser	Tyr	Asn	Leu	Asp	Phe	Leu	Glu	Arg	Trp	Glu	Asn	Leu
225					230					235					240
Gln	Lys	Lys	Lys	Ser	Pro	Ser	Tyr	Glu	Ser	Lys	Glu	Ala	Lys	Ser	Ala
				245				250					255		
Ile	Leu	Leu	Asn	Glu	Arg	Ala	Arg	Leu	Thr	Ala	Ala	His	Pro	Ile	Gln
			260					265					270		
Asp	Tyr	Ser	Leu	Met	Asn	Tyr	Ile	Ala	Tyr	Val	Thr	Tyr	Thr	Pro	Leu
	275						280					285			
Phe	Ile	Ala	Gly	Pro	Ile	Ile	Thr	Phe	Asn	Asp	Tyr	Val	Tyr	Gln	Ser
	290					295					300				
Lys	His	Thr	Leu	Pro	Ser	Ile	Asn	Phe	Lys	Phe	Ile	Phe	Tyr	Tyr	Ala
305					310					315					320
Val	Arg	Phe	Val	Ile	Ala	Leu	Leu	Ser	Met	Glu	Phe	Ile	Leu	His	Phe
				325				330					335		
Leu	His	Val	Val	Ala	Ile	Ser	Lys	Thr	Lys	Ala	Trp	Glu	Asn	Asp	Thr
			340					345					350		
Pro	Phe	Gln	Ile	Ser	Met	Ile	Gly	Leu	Phe	Asn	Leu	Asn	Ile	Ile	Trp
	355						360					365			
Leu	Lys	Leu	Leu	Ile	Pro	Trp	Arg	Leu	Phe	Arg	Leu	Trp	Ala	Leu	Leu
	370					375					380				
Asp	Gly	Ile	Asp	Thr	Pro	Glu	Asn	Met	Ile	Arg	Cys	Val	Asp	Asn	Asn
385					390					395					400

Tyr	Ser	Ser	Leu	Ala	Phe	Trp	Arg	Ala	Trp	His	Arg	Ser	Tyr	Asn	Lys
				405					410					415	
Trp	Val	Val	Arg	Tyr	Ile	Tyr	Ile	Pro	Leu	Gly	Gly	Ser	Lys	Asn	Arg
			420					425					430		
Val	Leu	Thr	Ser	Leu	Ala	Val	Phe	Ser	Phe	Val	Ala	Ile	Trp	His	Asp
		435					440					445			
Ile	Glu	Leu	Lys	Leu	Leu	Leu	Trp	Gly	Trp	Leu	Ile	Val	Leu	Phe	Leu
	450					455					460				
Leu	Pro	Glu	Ile	Phe	Ala	Thr	Gln	Ile	Phe	Ser	His	Tyr	Thr	Asp	Ala
465					470					475					480
Val	Trp	Tyr	Arg	His	Val	Cys	Ala	Val	Gly	Ala	Val	Phe	Asn	Ile	Trp
			485						490					495	
Val	Met	Met	Ile	Ala	Asn	Leu	Phe	Gly	Phe	Cys	Leu	Gly	Ser	Asp	Gly
			500					505					510		
Thr	Lys	Lys	Leu	Leu	Ser	Asp	Met	Phe	Cys	Thr	Val	Ser	Gly	Phe	Lys
		515					520					525			
Phe	Val	Ile	Leu	Ala	Ser	Val	Ser	Leu	Phe	Ile	Ala	Val	Gln	Ile	Met
	530					535					540				
Phe	Glu	Ile	Arg	Glu	Glu	Glu	Lys	Arg	His	Gly	Ile	Tyr	Leu	Lys	Cys
545					550					555					560

<210> 17
 <211> 609
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 17
 Met Ser Met Leu Arg Ile Trp Ser Cys Ile Val His Phe Phe Ser Val
 1 5 10 15
 Gln Ala Leu Asp Ser Arg Ile Lys Pro Asp Ile Glu Phe Lys Arg Arg
 20 25 30
 Gln Arg Ile Phe Ile Asn Ser Ser Lys Glu Glu Asn Gly Ser Ser Ser
 35 40 45
 Ser Ala Val Thr Val Thr Arg Asn Pro Val Leu Ser Ser Asn Ser Pro
 50 55 60
 Ser Pro Pro Leu Trp Asn Thr Trp Glu Phe Arg Leu Tyr Tyr Leu Ala
 65 70 75 80
 Phe Thr Val Val Val Pro Phe Met Ile Lys Ala Ala Leu Ala Thr Ser
 85 90 95
 Ser Glu Ser Asn Pro Asn Tyr Tyr Lys Phe Ser Gly Leu Leu Ala His
 100 105 110
 Gly Trp Ile Leu Gly Arg Lys Val Asp Asn Ser Asp Pro Gln Tyr Arg
 115 120 125
 Phe Phe Arg Ser Asn Phe Phe Leu Leu Ala Ile Leu Ile Leu Leu Gln
 130 135 140
 Ile Ile Leu Lys Lys Val Phe Val Lys Phe Ser Lys Ile Pro Lys Thr
 145 150 155 160
 Lys Phe Asp Phe Ala Cys Gly Leu Val Phe Val Cys Phe Met Tyr Gly
 165 170 175
 Ile Asn Ser Val Lys Leu Phe Thr His Ala Phe Ile Phe Phe Thr Leu
 180 185 190
 Ala His Ser Leu Lys Arg Lys Arg Leu Ile Ala Ala Phe Ala Ile Trp
 195 200 205
 Ser Tyr Gly Ile Phe Thr Leu Phe Ile Asn Gln Lys Met Lys Asn Leu
 210 215 220

Pro	Phe	Asn	Asn	Ile	Ala	Ile	Ile	Leu	Ser	Pro	Met	Asp	Gln	Trp	Tyr
225					230					235					240
Lys	Gly	Ile	Val	Pro	Arg	Trp	Asp	Phe	Phe	Phe	Asn	Phe	Thr	Leu	Leu
				245					250					255	
Arg	Leu	Leu	Ser	Tyr	Ser	Met	Asp	Phe	Leu	Glu	Arg	Trp	His	Glu	Gln
			260					265					270		
Leu	Ser	Arg	Gln	Pro	Ser	Ile	Asp	Tyr	Asp	Asp	Arg	Arg	Pro	Glu	Phe
		275					280				285				
Arg	Lys	Ser	Leu	Ser	Gly	Ser	Thr	Leu	Gln	Thr	Ile	Tyr	Glu	Ser	Gly
	290					295					300				
Lys	Asn	Val	Leu	Glu	Glu	Lys	Glu	Arg	Leu	Val	Ala	Glu	His	His	Ile
305					310					315					320
Gln	Asp	Tyr	Asn	Phe	Ile	Asn	Phe	Ile	Ala	Tyr	Ile	Thr	Tyr	Ala	Pro
			325						330					335	
Leu	Phe	Leu	Val	Gly	Pro	Ile	Ile	Thr	Phe	Asn	Asp	Tyr	Leu	Tyr	Gln
			340					345					350		
Ser	Glu	Asn	Lys	Leu	Pro	Ser	Leu	Thr	Lys	Lys	Asn	Ile	Gly	Phe	Tyr
		355					360					365			
Ala	Leu	Lys	Val	Phe	Ser	Ser	Leu	Leu	Leu	Met	Glu	Ile	Ile	Leu	His
	370					375					380				
Tyr	Ile	Tyr	Val	Gly	Ala	Ile	Ala	Arg	Thr	Lys	Ala	Trp	Asn	Asn	Asp
385					390					395					400
Thr	Pro	Leu	Gln	Gln	Ala	Met	Ile	Ala	Leu	Phe	Asn	Leu	Asn	Ile	Met
			405					410						415	
Tyr	Leu	Lys	Leu	Leu	Ile	Pro	Trp	Arg	Leu	Phe	Arg	Leu	Trp	Ala	Met
		420						425					430		
Val	Asp	Gly	Ile	Asp	Ala	Pro	Glu	Asn	Met	Leu	Arg	Cys	Val	Asp	Asn
	435						440					445			
Asn	Tyr	Ser	Thr	Val	Gly	Phe	Trp	Arg	Ala	Trp	His	Thr	Ser	Phe	Asn
	450					455					460				
Lys	Trp	Val	Ile	Arg	Tyr	Ile	Tyr	Val	Pro	Phe	Gly	Gly	Ser	Asn	Asn
465					470					475					480
Lys	Ile	Leu	Thr	Ser	Phe	Ala	Val	Phe	Ser	Phe	Val	Ala	Ile	Trp	His
			485					490						495	
Asp	Ile	Gln	Leu	Arg	Val	Leu	Phe	Trp	Gly	Trp	Leu	Thr	Val	Leu	Leu
		500						505					510		
Leu	Leu	Gly	Glu	Thr	Tyr	Ile	Thr	Asn	Cys	Phe	Ser	Arg	Tyr	Arg	Phe
		515					520					525			
Arg	Ser	Trp	Tyr	Arg	Phe	Val	Cys	Gly	Ile	Gly	Ala	Ala	Ile	Asn	Ile
	530					535					540				
Cys	Met	Met	Met	Ile	Ile	Asn	Val	Tyr	Gly	Phe	Cys	Leu	Gly	Ala	Glu
545					550					555					560
Gly	Thr	Lys	Leu	Leu	Leu	Lys	Gly	Ile	Phe	Asn	Asn	Ser	His	Ser	Pro
			565					570						575	
Glu	Phe	Leu	Thr	Ala	Val	Met	Val	Ser	Leu	Phe	Ile	Ala	Val	Gln	Val
			580					585					590		
Met	Phe	Glu	Ile	Arg	Glu	Glu	Glu	Lys	Arg	His	Gly	Ile	Asn	Leu	Lys
		595					600					605			

Cys

<210> 18

<211> 669

<212> PRT

<213> Saccharomyces cerevisiae

<400> 18

Met	Ser	Asn	Pro	Gln	Lys	Ala	Leu	Asn	Asp	Phe	Leu	Ser	Ser	Glu	Ser
1				5					10					15	
Val	His	Thr	His	Asp	Ser	Ser	Arg	Lys	Gln	Ser	Asn	Lys	Gln	Ser	Ser
			20					25					30		
Asp	Glu	Gly	Arg	Ser	Ser	Ser	Gln	Pro	Ser	His	His	His	Ser	Gly	Gly
		35					40					45			
Thr	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Ser	Asn	Asn
	50					55						60			
Asn	Asn	Asn	Gly	Asn	Asp	Gly	Gly	Asn	Asp	Asp	Asp	Tyr	Asp	Tyr	Glu
65					70					75					80
Met	Gln	Asp	Tyr	Arg	Pro	Ser	Pro	Gln	Ser	Ala	Arg	Pro	Thr	Pro	Thr
				85					90					95	
Tyr	Val	Pro	Gln	Tyr	Ser	Val	Glu	Ser	Gly	Thr	Ala	Phe	Pro	Ile	Gln
			100					105						110	
Glu	Val	Ile	Pro	Ser	Ala	Tyr	Ile	Asn	Thr	Gln	Asp	Ile	Asn	His	Lys
		115					120					125			
Asp	Asn	Gly	Pro	Pro	Ser	Ala	Ser	Ser	Asn	Arg	Ala	Phe	Arg	Pro	Arg
	130					135					140				
Gly	Gln	Thr	Thr	Val	Ser	Ala	Asn	Val	Leu	Asn	Ile	Glu	Asp	Phe	Tyr
145					150					155					160
Lys	Asn	Ala	Asp	Asp	Ala	His	Thr	Ile	Pro	Glu	Ser	His	Leu	Ser	Arg
				165					170					175	
Arg	Arg	Ser	Arg	Ser	Arg	Ala	Thr	Ser	Asn	Ala	Gly	His	Ser	Ala	Asn
			180					185					190		
Thr	Gly	Ala	Thr	Asn	Gly	Arg	Thr	Thr	Gly	Ala	Gln	Thr	Asn	Met	Glu
		195					200					205			
Ser	Asn	Glu	Ser	Pro	Arg	Asn	Val	Pro	Ile	Met	Val	Lys	Pro	Lys	Thr
	210					215					220				
Leu	Tyr	Gln	Asn	Pro	Gln	Thr	Pro	Thr	Val	Leu	Pro	Ser	Thr	Tyr	His
225					230					235					240
Pro	Ile	Asn	Lys	Trp	Ser	Ser	Val	Lys	Asn	Thr	Tyr	Leu	Lys	Glu	Phe
				245					250					255	
Leu	Ala	Glu	Phe	Met	Gly	Thr	Met	Val	Met	Ile	Ile	Phe	Gly	Ser	Ala
			260				265						270		
Val	Val	Cys	Gln	Val	Asn	Val	Ala	Gly	Lys	Ile	Gln	Gln	Asp	Asn	Phe
		275					280					285			
Asn	Val	Ala	Leu	Asp	Asn	Leu	Asn	Val	Thr	Gly	Ser	Ser	Ala	Glu	Thr
	290					295					300				
Ile	Asp	Ala	Met	Lys	Ser	Leu	Thr	Ser	Leu	Val	Ser	Ser	Val	Ala	Gly
305					310					315					320
Gly	Thr	Phe	Asp	Asp	Val	Ala	Leu	Gly	Trp	Ala	Ala	Ala	Val	Val	Met
			325						330					335	
Gly	Tyr	Phe	Cys	Ala	Gly	Gly	Ser	Ala	Ile	Ser	Gly	Ala	His	Leu	Asn
			340				345						350		
Pro	Ser	Ile	Thr	Leu	Ala	Asn	Leu	Val	Tyr	Arg	Gly	Phe	Pro	Leu	Lys
		355					360					365			
Lys	Val	Pro	Tyr	Tyr	Phe	Ala	Gly	Gln	Leu	Ile	Gly	Ala	Phe	Thr	Gly
	370					375					380				
Ala	Leu	Ile	Leu	Phe	Ile	Trp	Tyr	Lys	Arg	Val	Leu	Gln	Glu	Ala	Tyr
385					390					395					400
Ser	Asp	Trp	Trp	Met	Asn	Glu	Ser	Val	Ala	Gly	Met	Phe	Cys	Val	Phe
				405					410					415	
Pro	Lys	Pro	Tyr	Leu	Ser	Ser	Gly	Arg	Gln	Phe	Phe	Ser	Glu	Phe	Leu
			420					425					430		

Cys	Gly	Ala	Met	Leu	Gln	Ala	Gly	Thr	Phe	Ala	Leu	Thr	Asp	Pro	Tyr
	435						440					445			
Thr	Cys	Leu	Ser	Ser	Asp	Val	Phe	Pro	Leu	Met	Met	Phe	Ile	Leu	Ile
	450					455					460				
Phe	Ile	Ile	Asn	Ala	Ser	Met	Ala	Tyr	Gln	Thr	Gly	Thr	Ala	Met	Asn
465					470					475					480
Leu	Ala	Arg	Asp	Leu	Gly	Pro	Arg	Leu	Ala	Leu	Tyr	Ala	Val	Gly	Phe
			485						490					495	
Asp	His	Lys	Met	Leu	Trp	Val	His	His	His	His	Phe	Phe	Trp	Val	Pro
		500					505						510		
Met	Val	Gly	Pro	Phe	Ile	Gly	Ala	Leu	Met	Gly	Gly	Leu	Val	Tyr	Asp
	515					520						525			
Val	Cys	Ile	Tyr	Gln	Gly	His	Glu	Ser	Pro	Val	Asn	Trp	Ser	Leu	Pro
	530				535						540				
Val	Tyr	Lys	Glu	Met	Ile	Met	Arg	Ala	Trp	Phe	Arg	Arg	Pro	Gly	Trp
545					550					555					560
Lys	Lys	Arg	Asn	Arg	Ala	Arg	Arg	Thr	Ser	Asp	Leu	Ser	Asp	Phe	Ser
			565					570						575	
Tyr	Asn	Asn	Asp	Asp	Asp	Glu	Glu	Phe	Gly	Glu	Arg	Met	Ala	Leu	Gln
			580					585					590		
Lys	Thr	Lys	Thr	Lys	Ser	Ser	Ile	Ser	Asp	Asn	Glu	Asn	Glu	Ala	Gly
	595					600					605				
Glu	Lys	Lys	Val	Gln	Phe	Lys	Ser	Val	Gln	Arg	Gly	Lys	Arg	Thr	Phe
	610					615					620				
Gly	Gly	Ile	Pro	Thr	Ile	Leu	Glu	Glu	Glu	Asp	Ser	Ile	Glu	Thr	Ala
625					630					635					640
Ser	Leu	Gly	Ala	Thr	Thr	Thr	Asp	Ser	Ile	Gly	Leu	Ser	Asp	Thr	Ser
			645					650						655	
Ser	Glu	Asp	Ser	His	Tyr	Gly	Asn	Ala	Lys	Lys	Val	Thr			
		660						665							

<210> 19
 <211> 713
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 19
 Met Ser Pro Ser Ala Val Gln Ser Ser Lys Leu Glu Glu Gln Ser Ser
 1 5 10 15
 Glu Ile Asp Lys Leu Lys Ala Lys Met Ser Gln Ser Ala Ala Thr Ala
 20 25 30
 Gln Gln Lys Lys Glu His Glu Tyr Glu His Leu Thr Ser Val Lys Ile
 35 40 45
 Val Pro Gln Arg Pro Ile Ser Asp Arg Leu Gln Pro Ala Ile Ala Thr
 50 55 60
 His Tyr Ser Pro His Leu Asp Gly Leu Gln Asp Tyr Gln Arg Leu His
 65 70 75 80
 Lys Glu Ser Ile Glu Asp Pro Ala Lys Phe Phe Gly Ser Lys Ala Thr
 85 90 95
 Gln Phe Leu Asn Trp Ser Lys Pro Phe Asp Lys Val Phe Ile Pro Asp
 100 105 110
 Pro Lys Thr Gly Arg Pro Ser Phe Gln Asn Asn Ala Trp Phe Leu Asn
 115 120 125
 Gly Gln Leu Asn Ala Cys Tyr Asn Cys Val Asp Arg His Ala Leu Lys
 130 135 140

Thr	Pro	Asn	Lys	Lys	Ala	Ile	Ile	Phe	Glu	Gly	Asp	Glu	Pro	Gly	Gln
145					150					155					160
Gly	Tyr	Ser	Ile	Thr	Tyr	Lys	Glu	Leu	Leu	Glu	Glu	Val	Cys	Gln	Val
				165					170					175	
Ala	Gln	Val	Leu	Thr	Tyr	Ser	Met	Gly	Val	Arg	Lys	Gly	Asp	Thr	Val
			180					185					190		
Ala	Val	Tyr	Met	Pro	Met	Val	Pro	Glu	Ala	Ile	Ile	Thr	Leu	Leu	Ala
		195					200					205			
Ile	Ser	Arg	Ile	Gly	Ala	Ile	His	Ser	Val	Val	Phe	Ala	Gly	Phe	Ser
	210					215					220				
Ser	Asn	Ser	Leu	Arg	Asp	Arg	Ile	Asn	Asp	Gly	Asp	Ser	Lys	Val	Val
225					230					235					240
Ile	Thr	Thr	Asp	Glu	Ser	Asn	Arg	Gly	Gly	Lys	Val	Ile	Glu	Thr	Lys
			245						250					255	
Arg	Ile	Val	Asp	Asp	Ala	Leu	Arg	Glu	Thr	Pro	Gly	Val	Arg	His	Val
			260					265					270		
Leu	Val	Tyr	Arg	Lys	Thr	Asn	Asn	Pro	Ser	Val	Ala	Phe	His	Ala	Pro
		275					280					285			
Arg	Asp	Leu	Asp	Trp	Ala	Thr	Glu	Lys	Lys	Lys	Tyr	Lys	Thr	Tyr	Tyr
	290					295					300				
Pro	Cys	Thr	Pro	Val	Asp	Ser	Glu	Asp	Pro	Leu	Phe	Leu	Leu	Tyr	Thr
305					310					315					320
Ser	Gly	Ser	Thr	Gly	Ala	Pro	Lys	Gly	Val	Gln	His	Ser	Thr	Ala	Gly
				325					330					335	
Tyr	Leu	Leu	Gly	Ala	Leu	Leu	Thr	Met	Arg	Tyr	Thr	Phe	Asp	Thr	His
			340					345					350		
Gln	Glu	Asp	Val	Phe	Phe	Thr	Ala	Gly	Asp	Ile	Gly	Trp	Ile	Thr	Gly
		355					360					365			
His	Thr	Tyr	Val	Val	Tyr	Gly	Pro	Leu	Leu	Tyr	Gly	Cys	Ala	Thr	Leu
	370					375					380				
Val	Phe	Glu	Gly	Thr	Pro	Ala	Tyr	Pro	Asn	Tyr	Ser	Arg	Tyr	Trp	Asp
385					390					395					400
Ile	Ile	Asp	Glu	His	Lys	Val	Thr	Gln	Phe	Tyr	Val	Ala	Pro	Thr	Ala
				405					410					415	
Leu	Arg	Leu	Leu	Lys	Arg	Ala	Gly	Asp	Ser	Tyr	Ile	Glu	Asn	His	Ser
			420					425					430		
Leu	Lys	Ser	Leu	Arg	Cys	Leu	Gly	Ser	Val	Gly	Glu	Pro	Ile	Ala	Ala
		435					440					445			
Glu	Val	Trp	Glu	Trp	Tyr	Ser	Glu	Lys	Ile	Gly	Lys	Asn	Glu	Ile	Pro
	450					455					460				
Ile	Val	Asp	Thr	Tyr	Trp	Gln	Thr	Glu	Ser	Gly	Ser	His	Leu	Val	Thr
465					470					475					480
Pro	Leu	Ala	Gly	Gly	Val	Thr	Pro	Met	Lys	Pro	Gly	Ser	Ala	Ser	Phe
				485					490					495	
Pro	Phe	Phe	Gly	Ile	Asp	Ala	Val	Val	Leu	Asp	Pro	Asn	Thr	Gly	Glu
			500					505					510		
Glu	Leu	Asn	Thr	Ser	His	Ala	Glu	Gly	Val	Leu	Ala	Val	Lys	Ala	Ala
		515					520					525			
Trp	Pro	Ser	Phe	Ala	Arg	Thr	Ile	Trp	Lys	Asn	His	Asp	Arg	Tyr	Leu
	530					535					540				
Asp	Thr	Tyr	Leu	Asn	Pro	Tyr	Pro	Gly	Tyr	Tyr	Phe	Thr	Gly	Asp	Gly
545				550						555					560
Ala	Ala	Lys	Asp	Lys	Asp	Gly	Tyr	Ile	Trp	Ile	Leu	Gly	Arg	Val	Asp
				565					570					575	
Asp	Val	Val	Asn	Val	Ser	Gly	His	Arg	Leu	Ser	Thr	Ala	Glu	Ile	Glu
			580					585					590		

Ala Ala Ile Ile Glu Asp Pro Ile Val Ala Glu Cys Ala Val Val Gly
 595 600 605
 Phe Asn Asp Asp Leu Thr Gly Gln Ala Val Ala Ala Phe Val Val Leu
 610 615 620
 Lys Asn Lys Ser Ser Trp Ser Thr Ala Thr Asp Asp Glu Leu Gln Asp
 625 630 635 640
 Ile Lys Lys His Leu Val Phe Thr Val Arg Lys Asp Ile Gly Pro Phe
 645 650 655
 Ala Ala Pro Lys Leu Ile Ile Leu Val Asp Asp Leu Pro Lys Thr Arg
 660 665 670
 Ser Gly Lys Ile Met Arg Arg Ile Leu Arg Lys Ile Leu Ala Gly Glu
 675 680 685
 Ser Asp Gln Leu Gly Asp Val Ser Thr Leu Ser Asn Pro Gly Ile Val
 690 695 700
 Arg His Leu Ile Asp Ser Val Lys Leu
 705 710

<210> 20
 <211> 683
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 20
 Met Thr Ile Lys Glu His Lys Val Val Tyr Glu Ala His Asn Val Lys
 1 5 10 15
 Ala Leu Lys Ala Pro Gln His Phe Tyr Asn Ser Gln Pro Gly Lys Gly
 20 25 30
 Tyr Val Thr Asp Met Gln His Tyr Gln Glu Met Tyr Gln Gln Ser Ile
 35 40 45
 Asn Glu Pro Glu Lys Phe Phe Asp Lys Met Ala Lys Glu Tyr Leu His
 50 55 60
 Trp Asp Ala Pro Tyr Thr Lys Val Gln Ser Gly Ser Leu Asn Asn Gly
 65 70 75 80
 Asp Val Ala Trp Phe Leu Asn Gly Lys Leu Asn Ala Ser Tyr Asn Cys
 85 90 95
 Val Asp Arg His Ala Phe Ala Asn Pro Asp Lys Pro Ala Leu Ile Tyr
 100 105 110
 Glu Ala Asp Asp Glu Ser Asp Asn Lys Ile Ile Thr Phe Gly Glu Leu
 115 120 125
 Leu Arg Lys Val Ser Gln Ile Ala Gly Val Leu Lys Ser Trp Gly Val
 130 135 140
 Lys Lys Gly Asp Thr Val Ala Ile Tyr Leu Pro Met Ile Pro Glu Ala
 145 150 155 160
 Val Ile Ala Met Leu Ala Val Ala Arg Ile Gly Ala Ile His Ser Val
 165 170 175
 Val Phe Ala Gly Phe Ser Ala Gly Ser Leu Lys Asp Arg Val Val Asp
 180 185 190
 Ala Asn Ser Lys Val Val Ile Thr Cys Asp Glu Gly Lys Arg Gly Gly
 195 200 205
 Lys Thr Ile Asn Thr Lys Lys Ile Val Asp Glu Gly Leu Asn Gly Val
 210 215 220
 Asp Leu Val Ser Arg Ile Leu Val Phe Gln Arg Thr Gly Thr Glu Gly
 225 230 235 240
 Ile Pro Met Lys Ala Gly Arg Asp Tyr Trp Trp His Glu Glu Ala Ala
 245 250 255

Lys	Gln	Arg	Thr	Tyr	Leu	Pro	Pro	Val	Ser	Cys	Asp	Ala	Glu	Asp	Pro	260	265	270
Leu	Phe	Leu	Leu	Tyr	Thr	Ser	Gly	Ser	Thr	Gly	Ser	Pro	Lys	Gly	Val	275	280	285
Val	His	Thr	Thr	Gly	Gly	Tyr	Leu	Leu	Gly	Ala	Ala	Leu	Thr	Thr	Arg	290	295	300
Tyr	Val	Phe	Asp	Ile	His	Pro	Glu	Asp	Val	Leu	Phe	Thr	Ala	Gly	Asp	305	310	315
Val	Gly	Trp	Ile	Thr	Gly	His	Thr	Tyr	Ala	Leu	Tyr	Gly	Pro	Leu	Thr	325	330	335
Leu	Gly	Thr	Ala	Ser	Ile	Ile	Phe	Glu	Ser	Thr	Pro	Ala	Tyr	Pro	Asp	340	345	350
Tyr	Gly	Arg	Tyr	Trp	Arg	Ile	Ile	Gln	Arg	His	Lys	Ala	Thr	His	Phe	355	360	365
Tyr	Val	Ala	Pro	Thr	Ala	Leu	Arg	Leu	Ile	Lys	Arg	Val	Gly	Glu	Ala	370	375	380
Glu	Ile	Ala	Lys	Tyr	Asp	Thr	Ser	Ser	Leu	Arg	Val	Leu	Gly	Ser	Val	385	390	395
Gly	Glu	Pro	Ile	Ser	Pro	Asp	Leu	Trp	Glu	Trp	Tyr	His	Glu	Lys	Val	405	410	415
Gly	Asn	Lys	Asn	Cys	Val	Ile	Cys	Asp	Thr	Met	Trp	Gln	Thr	Glu	Ser	420	425	430
Gly	Ser	His	Leu	Ile	Ala	Pro	Leu	Ala	Gly	Ala	Val	Pro	Thr	Lys	Pro	435	440	445
Gly	Ser	Ala	Thr	Val	Pro	Phe	Phe	Gly	Ile	Asn	Ala	Cys	Ile	Ile	Asp	450	455	460
Pro	Val	Thr	Gly	Val	Glu	Leu	Glu	Gly	Asn	Asp	Val	Glu	Gly	Val	Leu	465	470	475
Ala	Val	Lys	Ser	Pro	Trp	Pro	Ser	Met	Ala	Arg	Ser	Val	Trp	Asn	His	485	490	495
His	Asp	Arg	Tyr	Met	Asp	Thr	Tyr	Leu	Lys	Pro	Tyr	Pro	Gly	His	Tyr	500	505	510
Phe	Thr	Gly	Asp	Gly	Ala	Gly	Arg	Asp	His	Asp	Gly	Tyr	Tyr	Trp	Ile	515	520	525
Arg	Gly	Arg	Val	Asp	Asp	Val	Val	Asn	Val	Ser	Gly	His	Arg	Leu	Ser	530	535	540
Thr	Ser	Glu	Ile	Glu	Ala	Ser	Ile	Ser	Asn	His	Glu	Asn	Val	Ser	Glu	545	550	555
Ala	Ala	Val	Val	Gly	Ile	Pro	Asp	Glu	Leu	Thr	Gly	Gln	Thr	Val	Val	565	570	575
Ala	Tyr	Val	Ser	Leu	Lys	Asp	Gly	Tyr	Leu	Gln	Asn	Asn	Ala	Thr	Glu	580	585	590
Gly	Asp	Ala	Glu	His	Ile	Thr	Pro	Asp	Asn	Leu	Arg	Arg	Glu	Leu	Ile	595	600	605
Leu	Gln	Val	Arg	Gly	Glu	Ile	Gly	Pro	Phe	Ala	Ser	Pro	Lys	Thr	Ile	610	615	620
Ile	Leu	Val	Arg	Asp	Leu	Pro	Arg	Thr	Arg	Ser	Gly	Lys	Ile	Met	Arg	625	630	635
Arg	Val	Leu	Arg	Lys	Val	Ala	Ser	Asn	Glu	Ala	Glu	Gln	Leu	Gly	Asp	645	650	655
Leu	Thr	Thr	Leu	Ala	Asn	Pro	Glu	Val	Val	Pro	Ala	Ile	Ile	Ser	Ala	660	665	670
Val	Glu	Asn	Gln	Phe	Phe	Ser	Gln	Lys	Lys	Lys						675	680	

Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Leu Phe Glu Ala Val Tyr
 405 410 415
 Gln Ile Val Tyr Asn Asn Val Arg Met Glu Asp Leu Pro Glu Met Ile
 420 425 430
 Glu Glu Leu Asp Ile Asp Asp Glu
 435 440

<210> 22
 <211> 391
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 22
 Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn
 1 5 10 15
 Ala Gly Arg Lys Arg Ser Ser Ser Ser Val Ser Leu Lys Ala Ala Glu
 20 25 30
 Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr
 35 40 45
 Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe
 50 55 60
 Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
 65 70 75 80
 Lys Leu Thr Glu Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr Leu
 85 90 95
 Pro Gly Ile Thr Leu Pro Asp Asn Leu Val Ala Asn Pro Asp Leu Ile
 100 105 110
 Asp Ser Val Lys Asp Val Asp Ile Ile Val Phe Asn Ile Pro His Gln
 115 120 125
 Phe Leu Pro Arg Ile Cys Ser Gln Leu Lys Gly His Val Asp Ser His
 130 135 140
 Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
 145 150 155 160
 Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
 165 170 175
 Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
 180 185 190
 Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
 195 200 205
 Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg
 210 215 220
 Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
 225 230 235 240
 Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
 245 250 255
 Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
 260 265 270
 Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
 275 280 285
 Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
 290 295 300
 Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
 305 310 315 320
 Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
 325 330 335

Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
 340 345 350
 Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
 355 360 365
 Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
 370 375 380
 Glu Leu Asp Leu His Glu Asp
 385 390

<210> 23
 <211> 2058
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> GPD2 disruption

<400> 23
 agatcttttg cggcgagggtg ccgatgggtt gctgagggga agagtgttta gcttacggac 60
 ctattgccat tgttattccg attaatctat tggtcagcag ctcttctcta ccctgtcatt 120
 ctagtatttt tttttttttt ttttggtttt actttttttt cttottgcct ttttttcttg 180
 ttactttttt tctagttttt tttccttcca ctaagctttt tccttgattt atccttgggt 240
 tcttctttct actcctttag attttttttt tatatatata tttttaagtt tatgtatttt 300
 ggtagattca attctctttc cctttccttt tccttcgctc cccttcctta tcaatgcttg 360
 ctgtcagaag attaacaaga tacacattcc ttaaggcctc gtccccgcgc ggtcaccgcg 420
 ccagcgacat ggaggcccag aataccctcc ttgacagtct tgacgtgcgc agctcagggg 480
 catgatgtga ctgtcgcccg tacatttagc ccatacatcc ccatgtataa tcatttgcat 540
 ccatacattt tgatggccgc acggcgcgaa gcaaaaatta cggtcctcgc ctgcagacct 600
 gcgagcaggg aaacgctccc ctcacagacg cgttgaattg tccccacgcc gcgcccctgt 660
 agagaaatat aaaagggttag gatttgccac tgaggttctt ctttcatata cttcctttta 720
 aaatcttgct aggatacagt tctcacatca catccgaaca taaacaacca tgtaaaatga 780
 ccactcttga cgacacggct taccggtagc gcaccagtgt cccggggggac gccgaggcca 840
 tcgaggcact ggatgggtcc ttcaccaccg acaccgtctt ccgcgtcacc gccaccgggg 900
 acggcttcac cctgcgggag gtgcgggtgg acccgcccct gaccaaggtg ttccccgacg 960
 acgaatcgga cgacgaatcg gacgccgggg aggacggcga cccggactcc cggacgttcg 1020
 tcgcgtacgg ggacgacggc gacctggcgg gcttcgtggt cgtctcgtac tccggctgga 1080
 accgccggct gaccgtcgag gacatcgagg tcgccccgga gcaccggggg cacgggggtc 1140
 ggcgcgcgtt gatggggctc gcgacggagt tcgcccgcga gcgggggcgc gggcacctct 1200
 ggctggaggt caccaacgtc aacgcaccgg cgatccacgc gtaccggcgg atggggttca 1260
 cctctgcgg cctggacacc gccctgtacg acggcaccgc ctcgacggc gagcaggcgc 1320
 tctacatgag catgccctgc ccctagtact gacaataaaa agattcttgt tttcaagaac 1380
 ttgtcatttg tatagttttt ttatattgta gttgttctat tttaatcaaa tgttagcgtg 1440
 atttatattt tttttcgcct cgacatcctc tgcccagatg cgaagttaag tgcgcagaaa 1500
 gtaatatcat gcgtcaatcg tatgtgaatg ctggtcgcta tactgctgtc gattcgatac 1560
 taacgccgcc atccagtgtc gacggatcct aggtgtacag ggcccaaaag ggogaattct 1620
 gcagatatcc atcacactgg cggccgctcg aggatagtct acaacaacgt ccgcatggaa 1680
 gacctaccgg agatgattga agagctagac atcgatgacg aatagacact ctcccccccc 1740
 ctccccctct gatctttcct gttgcctctt tttcccccaa ccaatttatc attatacaca 1800
 agttctacaa ctactactag taacattact acagttatta taattttcta ttctcttttt 1860
 ctttaagaat ctatcattaa cgttaatttc tatatataca taactaccat tatacacgct 1920
 attatcgttt acatatcaca tcaccgttaa tgaaagatac gacaccctgt aactaacac 1980
 aattaaataa tcgccataac cttttctgtt atctatagcc cttaaagctg tttcttcgag 2040
 ctttttcact gcagatct 2058

<210> 24
 <211> 816
 <212> DNA
 <213> *Saccharomyces cerevisiae*

<400> 24
 ccatatgatac atgtgtcgtc gcacacatat atatatgcct gtatgtgtca gcaactaaagt 60
 tgccctggcca tccacgctat atatacacgc ctggcgagat tgctcgagga ttgcctacgc 120
 gtgggcttga tccaccaacc aacgctcgcc aaatgaactg gcgctttggt cttctgccaat 180
 cgctccgtaaa ccccggccaa agagaccgga aagatcggtg aaaacatctt gatcttgctc 240
 ccgggaattt tagattcagg taggaaattg attacatcaa tactgttacc ctgaatcata 300
 ttcgacgatg tcgtctcaca cggaaatata attcatttct tggttttcca aaaaaatttt 360
 catttttttt cactttttttg ttctgtcctc cttttttttt tttttttttt attttttttc 420
 ctgtgttcac cttttttttt ttcagttgac atctttctgc attcttttct gtgttttttt 480
 tttttttttt cgtttttcca ttgttcgttc gttgcctggt ttttcgccct attgttctcg 540
 agcctaaaaa ttttttctt tctgtcttct ctttcttctg tcaaagtctt ctattccatt 600
 gttctctttg gtaaaactcat tgttgctgga actcagatat attcagggtc atttactgta 660
 cttcaattga cttttttctt gaaatttcaa cttgcctttt caacttggtc ttctttttta 720
 atcttattct acacttttagt tcccttacct tgttctaat tattgtctag caaaaagaaa 780
 acatacacct atttcattca cacactgcag aaaatg 816

<210> 25
 <211> 954
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> GCY1 PCR fragment

<400> 25
 ctgcagaaaa tgccctgctac ttacatgat tctacgaaaa tcccttctct aaatactgga 60
 gcccaaatac ctcaaataag tttaggtacg tggcagtcga aagagaacga tgcttataag 120
 gctgttttaa ccgctttgaa agatggctac cgacacattg atactgctgc tatttaccgt 180
 aatgaagacc aagtcggtca agccatcaag gattcagggt ttcctcgga agaaatcttt 240
 gttactacaa agttatggtg tacacaacac cacgaacctg aagtagcgt ggatcaatca 300
 ctaaagaggt taggattgga ctacgtagac ttatatgtga tgcattggcc tgccagatta 360
 gatccagcct acatcaaaaa tgaagacatc ttgagtgtgc caacaaagaa ggatggttct 420
 cgtgcagtgg atatcaccaa ttggaatttc atcaaaacct gggaattaat gcaggaaacta 480
 ccaaagactg gtaaaactaa ggccgttgga gtctccaaact tttctataaa taacctgaaa 540
 gatctattag catctcaagg taataagctt acgccaagct ctaaccaagt cgaaatacat 600
 ccattactac ctcaagaaga attgattaat ttttgtaaaa gtaaaggcat tgtggttgaa 660
 gcttattctc cgttaggtag taccgatgct ccactattga aggaaccggt tacccttgaa 720
 attgcgaaga aaaataacgt tcaaccgga cacgttggtt ttagctggca cgtccaaaga 780
 gggtatgttg tcttgccaaa atctgtgaat cccgatcgaa tcaaaacgaa caggaaaata 840
 tttactttgt ctactgagga ctttgaagct atcaataaca tatcgaagga aaagggcgaa 900
 aaaaggggtg tacatccaaa ttggtctcct ttccaagtat tcaagtaact taag 954

<210> 26
 <211> 2986
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> DAK1 PCR fragment

<400> 26

ctcgagtacc	ggtttcttct	tcagattccc	tcattggagaa	agtgcggcag	atgtatatga	60
cagagtcgcc	agtttccaag	agactttatt	caggcacttc	catgatagge	aagagagaag	120
acccagagat	gttggtgtcc	tagttacaca	tggtatttat	tccagagtat	tcctgatgaa	180
atggtttaga	tggacatacg	aagagtttga	atcgtttacc	aatglttccta	acgggagcgt	240
aatggtgatg	gaactggacg	aatccatcaa	tagatacgtc	ctgaggaccg	tgctacccaa	300
atggactgat	tgtgagggag	acctaactac	atagtgttta	aagattacgg	atatttaact	360
tacttagaat	aatgccattt	ttttgagtta	taataatcct	acgttagtgt	gagcgggatt	420
taaactgtga	ggacottaat	acattcagac	acttctgcgg	tatcaccccta	cttattccct	480
tcgagattat	atctaggaac	ccatcagggt	ggtggaagat	taccggttct	aagacttttc	540
agcttctctt	attgatgtta	cacctggaca	ccccttttct	ggcatccagt	ttttaatctt	600
cagtggcatg	tgagattctc	cgaaattaat	taaagcaatc	acacaattct	ctcggatacc	660
acctcggttg	aaactgacag	gtgggtttgt	acgcattgcta	atgcaaagga	gcctatatac	720
ctttggctcg	gctgctgtaa	caggggaatat	aaagggcagc	ataatttagg	agtttagtga	780
acttgcaaca	tttactattt	tccttcttta	cgtaaataatt	tttcttttta	attctaaatc	840
aatctttttc	aattttttgt	ttgtattctt	ttcttgctta	aatctataac	tacaaaaaac	900
acatacataa	atctagaaaa	tgtccgctaa	atcgtttgaa	gtcacagatc	cagtcaattc	960
aagtctcaaa	gggtttgccc	ttgctaacc	ctccattacg	ctggtccttg	aagaaaaaat	1020
tctcttcaga	aagaccgatt	ccgacaagat	cgcattaatt	tctggtggtg	gtagtggaca	1080
tgaacctaca	cacgcgggtt	tcattggtaa	gggtatgttg	agtggcgccg	tggttggcga	1140
aatttttgca	tccttctcaa	caaaacagat	tttaaattgca	atccgttttag	tcaatgaaaa	1200
tgctgtctgc	gttttattga	ttgtgaagaa	ctacacaggt	gatgttttgc	attttggtct	1260
gtccgctgag	agagcaagag	ccttggggtat	taactgccgc	gttgctgtca	taggtgatga	1320
tggtgcagtt	ggcagagaaa	agggtgggtat	ggttggtaga	agagcattgg	caggtaccgt	1380
tttggttcat	aagattgtag	gtgccttcgc	agaagaatat	tctagtaagt	atggcttaga	1440
cggtacagct	aaagtggcta	aaattatcaa	cgacaatttg	gtgaccattg	gatcttcttt	1500
agaccattgt	aaagtctctg	gcaggaaatt	cgaaagtga	ttaaacgaaa	aacaaatgga	1560
attgggtatg	ggtattcata	acgaacctgg	tgtgaaagtt	ttagacccta	ttccttctac	1620
cgaagacttg	atctccaagt	atatgctacc	aaaactattg	gatccaaacg	ataaggatag	1680
agcttttgta	aagtttgatg	aagatgatga	agttgtcttg	ttagttaaca	atctcggcgg	1740
tgtttctaata	tttgttatta	gttctatcac	ttccaaaact	acggatttct	taaaggaaaa	1800
ttacaacata	accccggttc	aaacaattgc	tggcacattg	atgacctcct	tcaatggtaa	1860
tgggttcagt	atcacattac	taaacgccac	taaggctaca	aaggctttgc	aatctgattt	1920
tgaggagatc	aaatcagtac	tagacttggt	gaacgcattt	acgaacgcac	cgggctggcc	1980
aattgcagat	tttgaaaaga	cttctgcccc	atctgttaac	gatgacttgt	tacataatga	2040
agtaacagca	aaggccgtcg	gtacctatga	ctttgacaag	tttgctgagt	ggatgaagag	2100
tggtgctgaa	caagttatca	agagcgaacc	gcacattacg	gaactagaca	atcaagttgg	2160
tgatggtgat	tgtggttaca	cttttagtggc	aggagttaaa	ggcatcaccg	aaaaccttga	2220
caagctgtcg	aaggactcat	tatctcaggc	ggttgcccac	atttcagatt	tcattgaagg	2280
ctcaatggga	ggtacttctg	gtggtttata	ttctattctt	ttgtcgggtt	tttcacacgg	2340
attaattcag	gtttgtaaat	caaaggatga	acccgtcact	aaggaaattg	tggctaagtc	2400
actcgggaatt	gcattggata	ctttatacaa	atatacaaag	gcaagggaagg	gatcatccac	2460
catgattgat	gctttagaac	cattcgttta	agaatttact	gcactaagg	atttcaataa	2520
ggcggtaaaa	gctgcagagg	aaggtgctaa	atccactgct	acattcgagg	ccaaatttgg	2580
cagagcttcg	tatgtcggcg	attcatctca	agtagaagat	cctggtgcag	taggcctatg	2640
tgagtttttg	aagggggttc	aaagcgcctt	gtaagtcgag	acaaatcgct	cttaaatata	2700
tacctaaaga	acattaaagc	tatattataa	gcaaagatac	gtaaattttg	cttatattat	2760
tatacacata	tcataattct	atatttttaa	gatttggtta	tataatgtac	glaatgcaaa	2820
ggaaataaat	tttatacatt	attgaacagc	gtccaagtaa	ctacattatg	tgactaata	2880
gtttagcgtc	gtgaagactt	tattgtgtcg	cgaaaagtaa	aaatttttaa	aattagagca	2940
ccttgaactt	gcgaaaaagg	ttctcatcaa	ctgttttaaa	cgtacg		2986

<210> 27

<211> 2651

<212> DNA

<213> Artificial Sequence

<220>

<223> GUP1 PCR fragment

<400> 27

```

ggtacctagg accggtttat cattatcaat actgccattt caaagaatac gtaaataatt      60
aatagtagtg attttcctaa ctttatttag tcaaaaaatt agccttttaa ttctgctgta     120
accggtacat gcccaaaata gggggcgggg tacacagaat atataacatc gtaggtgtct     180
gggtgaacag ttatttcctg gcatccacta aatataatgg agcccgtttt ttaagctggc     240
atccagaaaa aaaaagaatc ccagcaccaa aatattgttt tcttcaccaa ccatacgttc     300
atagggtccat tctcttagcg caactacaga gaacaggggc acaaacaggc aaaaaacggg     360
cacaacctca atggagtgat gcaacctgcc tggagtaaat gatgacacaa ggcaattgac     420
ccacgcatgt atctatctca ttttcttaca ccttctatta cttctgctc tctctgattt     480
ggaaaaagct gaaaaaaaag gttgaaacca gttccctgaa attattcccc tacttgacta     540
ataagtatat aaagacggta ggtattgatt gtaattctgt aaatctattt cttaaacttc     600
ttaaattcta cttttatagt tagtcttttt tttagtttta aaacaccaag aacttagttt     660
cgaataaaca cacataaaga attcgaaaat gtcgctgate agcatcctgt ctcccctaatt     720
tacttccgag ggcttagatt caagaatcaa accttcacca aaaaaggatg cctctactac     780
cactaagcca tcaactatgga aaactactga gttcaaattc tactacattg catttctggt     840
cgtgggtccc ttgatgttct atgctgggtt acaagctagt tcgcccgaaa atccaaacta     900
tgcaagatac gaacgtctcc tatctcaagg ttggttattt ggcagaaaag tagacaatag     960
tgattctcaa tatagggtttt tcagggacaa ttttgcgcta ttgtcagttt taatgctagt    1020
ccacacttct ataaaacgca ttgtacttta ttcaacaaat atcactaaat tgaggtttga    1080
tctgatattt ggtttgatct ttttagtggt cgctcatggt gtcaattcga taagaatttt    1140
agcccatatg ctaattttat atgccatcgc ccattgtact aagaacttta gaagaatagc    1200
caccatcagc atttggattt atggtatttc tacgcttttt attaacgaca acttcagagc    1260
atatccattt ggtaatatth gctctttttt aagcccattg gaccattggt atagagggtat    1320
cattccaaga tgggatgtct ttttcaattt tactcttttg agagtcttaa gttacaactt    1380
ggactttcta gagagggtgg agaatttaca aaagaagaaa agtccatcct atgaatcaaa    1440
agaagctaaa tcagccattt tgcctcaatga acgtgctaga ttaactgctg cacaccccat    1500
acaggactac agcttaatga atttatattgc atatgttact tacacgccac ttttcattgc    1560
cggccccatt ataacattca atgattatgt ttaccaatcg aaacatacct tgccatcaat    1620
aaatttcaaa ttcatttttt actatgcggg gagattcggt attgctctct tatctatgga    1680
gttcattttt cactttctcc acgttggtgg aatctcaaaa accaaagcgt gggaaaatga    1740
cacaccttcc cagatttcca tgattggctt atttaatttg aatattattt ggctaaaact    1800
actgattccg tggaggctgt ttaggctgtg ggctttgcta gacggaatcg atacacctga    1860
aaatatgata aggtgtgttg ataacaatta cagttcacta gcattctgga gagcttggca    1920
tagaagctac aataagtggg ttgtccgtta catatatatt cctctaggtg gttcaaaaaa    1980
tagagttttg acatcactag cagtcttttc cttcgtagct atatggcatg acatcgaact    2040
aaagttatta ttatgggggt ggctaatagt tttgttcttc ttaccagaaa tttttgctac    2100
ccaaattttc tctcattata ccgacgcagt ctggtacaga cacgtttgct ctgtcgggtgc    2160
tgttttcaac atatgggtta tgatgatcgc taatcttttt ggattctgct tgggctctga    2220
cggtactaaa aaattactaa gcgatatgtt ctgtaccgta tctgggttca aatttgtaat    2280
tttggaacag gttagtttat tcatcgcagt acaaataatg tttgaaatca gagaagaaga    2340
aaagaggcac ggaatttacc taaaatgctg aggatccctt tttcctttgt cgatatcatg    2400
taattagtta tgtcacgctt acattcacgc cctcctccca catccgctct aaccgaaaag    2460
gaaggagtta gacaacctga agtctagggt cctattttatt ttttttaata gttatgttag    2520
tattaagaac gttattttata tttcaaattt tttttttttt tctgtacaaa cgcggtgtacg    2580
catgtaacat tatactgaaa accttgcttg agaagggttt gggacgctcg aaggcttcct    2640
aggctcgagt t                                     2651

```

<210> 28

<211> 3037

<212> DNA

<213> Artificial Sequence

<220>

<223> FPS1 PCR fragment

<400> 28

aagagctccg	gactagtcgt	acgaattcta	tccttttggt	gtttccgggt	gtacaatatg	60
gacttcctct	tttctggcaa	ccaaacccat	acatcgggat	tcctataata	cottcgttgg	120
tctccctaac	atgtaggtgg	cggaggggag	atatacaata	gaacagatac	cagacaagac	180
ataatgggct	aaacaagact	acaccaatta	cactgcctca	ttgatgggtg	tacataacga	240
actaatactg	tagccctaga	cttgatagcc	atcatcatat	cgaagtttca	ctaccctttt	300
tccatttgcc	atctattgaa	gtaataatag	gcgcattgca	cttcttttct	ttttttttct	360
tttctctctc	ccccgttggt	gtctcaccat	atccgcaatg	acaaaaaat	gatggaagac	420
actaaaggaa	aaaattaacg	acaaagacag	caccaacaga	tgtcgttggt	ccagagctga	480
tgaggggtat	ctcgaagcac	acgaaacttt	ttccttcctt	cattcacgca	cactactctc	540
taatgagcaa	cggatacgg	ccttccttcc	agttacttga	atttgaaata	aaaaaagttt	600
gctgtcttgc	tatcaagtat	aaatagacct	gcaattatta	atcttttggt	tcctcgtcat	660
tgttctcggt	ccctttcttc	cttgtttctt	tttctgcaca	atatttcaag	ctataccaag	720
catacaatca	actccagctg	cattaaaatg	agtaatcctc	aaaaagctct	aaacgacttt	780
ctgtccagtg	aatctgttca	tacacatgat	agttctagga	aacaatctaa	taagcagtca	840
tccgacgaag	gacgtcttcc	atcacaacct	tcacatcctc	actctgggtg	tactaacaac	900
aataataaca	ataataataa	taataataac	agtaacaaca	acaacaacgg	caacgatggg	960
ggaaatgatg	acgactatga	ttatgaaatg	caagattata	gaccttctcc	gcaaagtgcg	1020
cggcctactc	ccacgtatgt	tcacacaatat	tctgtagaaa	gtgggactgc	tttcccgttt	1080
caagagggtta	ttcctagcgc	atacattaac	acacaagata	taaaccataa	agataacggt	1140
ccgcgcagtg	caagcagtaa	tagagcattc	aggcctagag	ggcagaccac	agtgtcggcc	1200
aacgtgctta	acattgaaga	tttttacaaa	aatgcagacg	atgcgcatac	catcccggag	1260
tcacatttat	cgagaaggag	aagtaggtcg	agggtctacg	gtaatgctgg	gcacagtgcc	1320
aatacaggcg	ccacgaatgg	caggactact	ggtgcccaaa	ctaatatgga	aagcaatgaa	1380
tcaccacgta	acgtccccat	tatggtgaag	ccaaagacat	tataccagaa	ccctcaaaca	1440
cctacagtct	tgcctccac	ataccatcca	attaataaat	ggtcttccgt	caaaaacact	1500
tatttgaaagg	aatttttagc	cgagtttatg	ggaacaatgg	ttatgattat	tttcggtagt	1560
gctgttggtt	gtcaggtcaa	tgttgctggg	aaaatacagc	aggacaattt	caacgtggct	1620
ttggataacc	ttaacgttac	cgggtcttct	gcagaaacga	tagacgctat	gaagagttta	1680
acatccttgg	tttcatccgt	tgcggggcgt	acctttgatg	atgtggcatt	gggctgggct	1740
gctgccgtgg	tgatgggcta	tttctgcgct	ggtggtagtg	ccatctcagg	tgctcathtt	1800
aatccgtcta	ttacattagc	caatttggtg	tatagagggt	ttcccctgaa	gaaagttcct	1860
tattactttg	ctggacaatt	gatcgggtgcc	ttcacaggcg	ctttgatctt	gtttathttgg	1920
tacaaaaggg	tggtacaaga	ggcatatagc	gattgggtgga	tgaatgaaag	tggtgcggga	1980
atgttttgcg	tttttccaaa	gccttatcta	agttcaggac	ggcaattttt	ttccgaattt	2040
ttatgtggag	ctatgttaca	agcaggaaca	tttgcgctga	ccgatcctta	tacgtgtttg	2100
tcctctgatg	ttttcccat	gatgatgttt	attttgattt	tcattatcaa	tgcttccatg	2160
gcttatcaga	caggtacagc	aatgaatttg	gctcgtgac	tgggcccaag	tcttgacta	2220
tatgcagttg	gatttgatca	taaaatgctt	tgggtgcac	atcatcattt	cttttgggtt	2280
cccatggtag	gcccatttat	tgggtgcgtta	atgggggggt	tggtttacga	tgtctgtatt	2340
tatcagggtc	atgaatctcc	agtcaactgg	tctttaccag	tttataagga	aatgattatg	2400
agagcctggg	ttagaaggcc	tgggttgaag	aagagaaata	gagcaagaag	aacatcggac	2460
ctgagtgact	tctcatacaa	taacgatgat	gatgaggaat	ttggagaaag	aatggctctt	2520
caaaagacaa	agaccaagtc	atctattttca	gacaacgaaa	atgaagcagg	agaaaagaaa	2580
gtgcaattta	aatctgttca	gcgcggcaaa	agaacgtttg	gtggtatacc	aacaattctt	2640
gaagaagaag	attccattga	aactgcttcg	ctaggtgcga	cgacgactga	ttctattggg	2700
ttatccgaca	catcatcaga	agattcgcac	tatggtaatg	ctaagaaggt	aacatgagga	2760
tccccttttc	ctttgtcgat	atcatgtaat	tagttatgtc	acgcttacat	tcacgccctc	2820
ctcccacatc	cgctctaacc	gaaaagggaag	gagttagaca	acctgaagtc	taggtcccta	2880
tttatTTTTT	ttaatagtta	tgttagtatt	aagaacgtta	tttatatttc	aaatttttct	2940

tttttttctg	tacaaacgcg	tgtacgcatg	taacattata	ctgaaaacct	tgcttgagaa	3000
ggttttggga	cgctcgaagg	cttcctaggg	tcgagtt			3037

<210> 29
 <211> 12954
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pRN605 plasmid for overexpression of GCY1, DAK1 and GUP1

<400> 29

tacgccaagc	tcggaattaa	ccctcactaa	agggaaacaaa	agctgggtac	cgggcccccc	60
ctcgagccta	ggaagccttc	gagcgtccca	aaaccttctc	aagcaagggt	ttcagtataa	120
tgttacatgc	gtacacgcgt	ttgtacagaa	aaaaaagaaa	aatttgaaat	ataaataacg	180
ttcttaatac	taacataact	attaaaaaaa	ataaataggg	acctagactt	caggttgtct	240
aactccttcc	ttttcgggta	gagcggatgt	gggaggaggg	cgtgaatgta	agcgtgacat	300
aactaattac	atgatatcga	caaaggaaaa	ggggatcctc	agcatttttag	gtaaattccg	360
tgcttctttt	cttcttctct	gatttcaaac	attatttgta	ctgcgatgaa	taaactaacg	420
cttgccaaaa	ttacaaattt	gaaaccagat	acggtacaga	acatatcgct	tagtaatttt	480
ttagtaccgt	cagagcccaa	gcagaatcca	aaaagattag	cgatcatcat	aacctatg	540
ttgaaaacag	caccgacagc	gcaaacgtgt	ctgtaccaga	ctgcgtcggg	ataatgagag	600
aaaatttggg	tagcaaaaat	ttctggtaag	aggaacaaaa	ctattagcca	accccataat	660
aataacttta	gttcgatgtc	atgccatata	gctacgaagg	aaaagactgc	tagtgatgtc	720
aaaactctat	tttttgaacc	acctagagga	atatatatgt	aacggacaac	ccacttattg	780
tagcttctat	gccaaactct	ccagaatgct	agtgaactgt	aattgttatc	aacacacctg	840
atcatatttt	cagggtgtatc	gattccgtct	agcaaagccc	acagcctaaa	cagcctccac	900
ggaatcagta	gttttagcca	aataatatcc	aaattaaata	agccaatcat	ggaaatctgg	960
aaaggtgtgt	cattttccca	cgctttgggt	tttgagattg	ccacaacgtg	gagaaagtgt	1020
aaaatgaact	ccatagataa	gagagcaata	acgaatctca	ccgcatagta	aaaaatgaat	1080
ttgaaattta	ttgatggcaa	ggtatgtttc	gattggtaaa	cataatcatt	gaatgttata	1140
atggggccgg	caatgaaaag	tggcgtgtaa	gtaacatatg	caatataatt	cattaagctg	1200
tagtccctgta	tggggtgtgc	agcagttaat	ctagcacgtt	cattgagcaa	aatggctgat	1260
ttagcttctt	ttgattcata	ggatggactt	ttcttctttt	gtaaattctc	ccacctctct	1320
aagaagtcca	agttgttaact	taagactctc	aaaagagtaa	aattgaaaaa	gacatcccat	1380
cttggaatga	tacctctata	ccaatgggtc	aatgggctta	aaaaagagca	aatattacca	1440
aatggatatg	ctctgaagtt	gtcgttaata	aaaagcgtag	aaataccata	aatccaaatg	1500
ctgatgggtg	ctattcttct	aaagttcttt	agtlacatggg	cgatggcata	taaaattagc	1560
atatgggcta	aaattcttat	cgaattgaca	ccatgagcgg	ccactaaaaa	gatcaaacca	1620
aatatcagat	caaacctcaa	tttagtgata	tttgttgaat	aaagtacaat	gcgttttata	1680
gaagtgtgga	ctagcattaa	aactgacaat	agcgcaaaaat	tgtccctgaa	aaacctatat	1740
tgagaatcac	tattgtctac	ttttctgcca	aataaccaac	cttgagatag	gagacgttcg	1800
tatcttgcac	agtttggatt	ttcgggcgaa	ctagcttgta	acctagcata	gaacatcaag	1860
ggaaccacga	ccagaaatgc	aatgtagtag	aatttgaaact	cagtagtttt	ccatagtgat	1920
ggcttagtg	tagtagaggc	atcctttttt	ggtgaagggt	tgattcttga	atctaagccc	1980
tcggaagtaa	ttagggggaga	caggatgctg	atcagcgaca	ttttgaattc	tttatgtgtg	2040
tttattcgaa	actaagttct	tgggtgtttta	aaactaaaaa	aaagactaac	tataaaaagta	2100
gaatttaaga	agtttaagaa	atagattttac	agaattacaa	tcaataccta	ccgtctttat	2160
atacttatta	gtcaagtagg	ggaataattt	caggggaactg	gtttcaacct	tttttttcag	2220
ctttttccaa	atcagagaga	gcagaaggta	atagaagggtg	taagaaaatg	agatagatac	2280
atgcgtgggt	caattgcctt	gtgtcatcat	ttactccagg	caggttgcat	cactccattg	2340
aggttgtgcc	cgttttttgc	ctgtttgtgc	ccctgttctc	tgtagttgcg	ctaagagaat	2400
ggacctatga	actgatgggt	ggtgaagaaa	acaatatattt	ggtgctggga	ttcttttttt	2460
ttctggatgc	cagcttaaaa	agcgggctcc	attatatatta	gtggatgcca	ggaataaaact	2520
gttcacccag	acacctacga	tgttatatat	tctgtgtaac	ccgcccccta	ttttgggcat	2580

gtacgggtta	cagcagaatt	aaaaggctaa	ttttttgact	aaataaagtt	aggaaaatca	2640
ctactattaa	ttattttacgt	attcttttgaa	atggcagtat	tgataatgat	aaaccggttt	2700
cttcttcaga	ttccctcatg	gagaaagtgc	ggcagatgta	tatgacagag	tcgccagttt	2760
ccaagagact	ttatttcaggc	acttccatga	taggcaagag	agaagaccca	gagatgttgt	2820
tgtectagtt	acacatggta	tttattccag	agtattcctg	atgaaatggg	ttagatggac	2880
atacgaagag	tttgaatcgt	ttaccaatgt	tcctaacggg	agcgtaatgg	tgatggaact	2940
ggacgaatcc	atcaatagat	acgtccctgag	gaccgtgcta	cccaaagga	ctgattgtga	3000
gggagaccta	actacatagt	gtttaaagat	tacggatatt	taacttactt	agaataatgc	3060
catttttttg	agttataata	atcctacgtt	agtgtgagcg	ggattttaa	tgtgaggacc	3120
ttaatacatt	cagacacttc	tgccgtatca	ccctacttat	tccttcgag	attatatcta	3180
ggaacccatc	aggttggtgg	aagattaccc	gttctaagac	ttttcagctt	cctctattga	3240
tgttacacct	ggacacccct	tttctggcat	ccagttttta	atcttcagtg	gcatgtgaga	3300
ttctccgaaa	ttaattaaag	caatcacaca	attctctcgg	ataccacctc	ggttgaaact	3360
gacaggtggg	ttgttacgca	tgctaataca	aaggagccta	tatacctttg	gctcggtgc	3420
tgtaacaggg	aatataaagg	gcagcataat	ttaggagttt	agtgaacttg	caacatttac	3480
tattttccct	tcttacgtaa	atatttttct	ttttaattct	aaatcaatct	ttttcaattt	3540
tttgtttgta	ttcttttctt	gcttaaattct	ataactacaa	aaaacacata	cataaatcta	3600
gaaaatgtcc	gctaaatcgt	ttgaagtcac	agatccagtc	aattcaagtc	tcaaaggggt	3660
tgcccttgct	aacccctcca	ttacgctggg	cctgaagaa	aaaattctct	tcagaaagac	3720
cgattccgac	aagatcgcat	taattttctg	tggtggtagt	ggacatgaac	ctacacacgc	3780
cggtttcatt	ggtaagggtg	tggtgagtg	cgccgtgggt	ggcgaaattt	ttgcatcccc	3840
ttcaacaaaa	cagattttta	atgcaatccg	tttagtcaat	gaaaatgctg	ctggcgtttt	3900
attgattgtg	aagaactaca	caggtgatgt	tttgcaattt	ggtctgtccg	ctgagagagc	3960
aagagccttg	ggtattaact	gccgcgttgc	tgatcatagg	gatgatgttg	cagttggcag	4020
agaaaagggt	ggtatgggtg	gtagaagagc	attggcaggt	accgttttgg	ttcataagat	4080
tgtaggtgcc	ttcgcagaag	aatattctag	taagtatggc	ttagacggta	cagctaaagt	4140
ggctaaaatt	atcaacgaca	atttggtgac	cattggatct	tcttttagacc	attgtaaagt	4200
tcctggcagg	aaattcgaaa	gtgaattaaa	cgaaaaacaa	atggaattgg	gtatgggtat	4260
tcataacgaa	cctggtgtga	aagtttttag	ccctattcct	tctaccgaag	acttgatctc	4320
caagtatatg	ctaccaaacc	tattggatcc	aaacgataag	gatagagctt	ttgtaaagtt	4380
tgatgaagat	gatgaagttg	tcttggttag	taacaatctc	ggcggtgttt	ctaattttgt	4440
tattagttct	atcacttcca	aaactacgga	tttcttaaag	gaaaattaca	acataacccc	4500
ggttcaaaca	attgctggca	cattgatgac	ctccttcaat	ggtaatgggt	tcagtatcac	4560
attactaaac	gccactaagg	ctacaaaggc	tttgcaatct	gatttttgagg	agatcaaate	4620
agtactagac	ttgttgaacg	catttacgaa	cgcaccgggc	tgcccaattg	cagattttga	4680
aaagacttct	gccccatctg	ttaacgatga	cttggtacat	aatgaagtaa	cagcaaaggc	4740
cgtcggtacc	tatgactttg	acaagtttgc	tgagtggatg	aagagtgggt	ctgaacaagt	4800
tatcaagagc	gaaccgcaca	ttacggaact	agacaatcaa	gttggtgatg	gtgattgtgg	4860
ttacacttta	gtggcaggag	ttaaaggcat	caccgaaaac	cttgacaagc	tgtcgaagga	4920
ctcattatct	caggcggttg	cccaaatttc	agatttcatt	gaaggctcaa	tgggaggtac	4980
ttctggtggg	ttatatctta	ttcttttgtc	gggtttttca	caaggattaa	ttcagggttg	5040
taaatcaaag	gatgaaccgc	tcactaagga	aattgtggct	aagtcactcg	gaattgcatt	5100
ggatacttta	tacaaatata	caaaggcaag	gaagggatca	tccaccatga	ttgatgcttt	5160
agaaccattc	gttaaagaat	ttactgcate	taaggatttc	aataaggcgg	taaaagctgc	5220
agaggaagg	gctaaatcca	ctgctacatt	cgaggccaaa	tttggcagag	cttcgtatgt	5280
cggcgattca	tctcaagtag	aagatcctgg	tgcagtaggc	ctatgtgagt	ttttgaagg	5340
ggttcaaagc	gccttgtaag	tcgagacaaa	tcgctcttaa	atatatacct	aaagaacatt	5400
aaagctatat	tataagcaaa	gatacgtaaa	ttttgcttat	attattatac	acatatcata	5460
tttctatatt	tttaagattt	ggttatataa	tgtacgtaat	gcaaaggaaa	taaattttat	5520
acattattga	acagcgtcca	agtaactaca	ttatgtgcac	taatagttaa	gcgtcgtgaa	5580
gactttattg	tgctcgcaaa	agtaaaaatt	ttaaaaatta	gagcaccttg	aacttgcgaa	5640
aaagggtctc	atcaactgtt	taaaacgtac	gtgtggaaga	acgattacaa	caggtgttgt	5700
cctctgagga	cataaaatac	acaccgagat	tcacaaactc	attgctggag	ttagcatatc	5760
tacaattggg	tgaaatgggg	agcgatttgc	aggcatttgc	tcggcatgcc	ggtagagggt	5820
tggtcaataa	gagcgacctc	atgctatacc	tgagaaagca	acctgacctc	caggaaagag	5880
ttactcaaga	ataagaattt	tcgttttaaa	acctaaagag	cacttttaaa	tttgtataca	5940

cttatttttt	ttataactta	tttaataata	aaaatcataa	atcataagaa	attcgcgcgc	6000
ttacttgaat	acttcgaaag	gagaccaatt	tggatgtaca	accctttttt	cgcccttttc	6060
cttcgatatg	ttattgatag	cttcaaagtc	ctcagtagac	aaagtaaata	ttttcctggt	6120
cgttttgatt	cgatcgggat	tcacagatth	tggcaagaca	acataacctc	tttggacgtg	6180
ccagctaata	acaacgtgtc	cggtttgaac	gttatttttc	ttcgcaattt	caaggataac	6240
cggttccttc	aatagtggag	catcggtact	acctaacgga	gaataagctt	caaccacaat	6300
gcctttactt	ttacaaaaat	taatcaattc	gtcttgaggt	agtaatggat	gtatttcgac	6360
ttggtttagca	gctggcgtaa	gcttattacc	ttgagatgct	aatagatctt	tcagggttatt	6420
tatagaaaag	ttggagactc	caacggcctt	agttttacca	gtctttggta	gttcctgcat	6480
taattcccag	gttttgatga	aattccaatt	ggtgatatac	actgcacgag	aaccatcctt	6540
ctttgttggc	acactcaaga	tgtcttcatt	tttgatgtag	gctggatcta	atctggcagg	6600
ccaatgcac	aaatataagt	ctacgtagtc	caatcctaac	ctcttttagtg	attgatccag	6660
cgtacttca	ggttcgtggt	gttgtgtaca	ccataacttt	gtagtaacaa	agattttctt	6720
ccgaggaaca	cctgaatcct	tgatggcctg	accgacttgg	tcttcattac	ggtaaatagc	6780
agcagtatca	atgtgtcggg	agccatcttt	caaagcgggt	aaaacagcct	tataagcatc	6840
gttctctttc	gactgccacg	tacctaaacc	tatttgaggg	atltgggctc	cagtatttag	6900
agaaaggatt	ttcgtagaat	catgtaaagt	agcaggcatt	ttctgcagtt	aattcagtaa	6960
atlttcgac	ttgggaagaa	aaaagcagta	agcgtgaaaa	atctaaaagc	tgatgtagta	7020
gaagatccta	ttctttaaca	aagattgacc	ttttcttttt	cttcttggtt	tgagtagaaa	7080
ggggaaggaa	gaatacaaga	gagaggaaaa	aaaggaaagt	aaaaagagag	cgtgatataa	7140
atgaatatat	attaaacaag	agagattggg	aaggaaagga	tcaaaacaaac	ccaaaaatat	7200
ttcaaaaagg	agagagagag	gcgagtttgg	tttcaaaacg	gtttatttat	ttatgcaaga	7260
ggacgtggaa	gaaaaagaag	aagggaagaa	aaaatttgaa	agaaaaaaac	gcgtggcggg	7320
taaagaagaa	aatggaaaat	agaggccggg	tgacagagaa	atattgaggg	ttatttgga	7380
aatatgttag	ggtgagggat	atgtttttta	gggttttgag	gatccgataa	ggaagaatgt	7440
aggttaaatg	ttgtgcatta	attgctgtgg	cagcttacc	gcttccccac	acatttacta	7500
gttctagagc	ggccgccacc	gcggtggagc	tccaattcgc	cctatagtga	gtcgtattac	7560
aattcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	tacccaactt	7620
aatgccttg	cagcacatcc	ccccttcgcc	agctggcgta	atagcgaaga	ggcccgcacc	7680
gacgcctct	cccaacagtt	gcgcagcctg	aatggcgaat	ggcgcgacgc	gccctgtagc	7740
ggcgcattaa	gcgcggcggg	tgtggtggtt	acgcgcagcg	tgaccgctac	acttgccagc	7800
gccctagcgc	ccgtcctttt	cgtttcttct	ccttcccttc	tcgccacggt	cgccggtagt	7860
gttagacctg	aacaagggtt	actaaaaatc	cgtaaagaac	ttcaattgta	cgccaactta	7920
aggcctcgtc	cccgcggggt	caccgcggcca	gcgacatgga	ggcccagaat	accctccttg	7980
acagtcttga	cgtgcgcagc	tcaggggcat	gatgtgactg	tcgcccgctac	atttagccca	8040
tacatcccca	tgtataatca	tttgcatcca	tacattttga	tgccgcacgc	gcgcgaagca	8100
aaaattacgg	ctcctcgtct	cagacctgcg	agcagggaaa	cgtcccccct	acagacgcgt	8160
gaattgtccc	cacgcgcgcg	ccctgtagag	aaatataaaa	ggttaggatt	tgccactgag	8220
gttcttcttt	catatacttc	cttttaaaat	cttgctagga	tacagttctc	acatcacatc	8280
cgaacataaa	caaccatggg	taaaaagcct	gaactcaccg	cgacgtctgt	cgagaagtth	8340
ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	tctcggaggg	cgaagaatct	8400
cgtgctttca	gcttcgatgt	aggaggcggt	ggatatgtcc	tgccgggtaaa	tagctgcgcc	8460
gatggtttct	acaaagatcg	ttatgtttat	cggcactttg	catcggcgcg	gctcccgatt	8520
ccggaagtgc	ttgacattgg	ggaattcagc	gagagcctga	cctattgcat	ctcccgcgt	8580
gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	tgcccgcgtg	tctgcagccg	8640
gtcgcggagg	ccatggatgc	gatcgcgtgc	gccgatctta	gccagacgag	cgggttcggc	8700
ccattcggac	cgcaagggaat	cggtcataac	actacatggc	gtgatttcat	atgcgcgatt	8760
gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	acaccgtcag	tgcgctccgtc	8820
gcgcaggctc	tcgatgagct	gatgcttttg	gccgaggact	gccccgaagt	ccggcacctc	8880
gtgcacgcgg	atlttcggctc	caacaatgtc	ctgacggaca	atggccgcat	aacagcggtc	8940
attgactgga	gcgaggcgat	gttcggggat	tcccaataacg	aggtcgccaa	catcttcttc	9000
tggaggccgt	ggttggtctg	tatggagcag	cagacgcgct	acttcgagcg	gaggcatccg	9060
gagcttgcag	gatcgccgcg	gctccggggc	tatatgctcc	gcattgggtct	tgaccaactc	9120
tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	gggcgcaggg	tcgatgcgac	9180
gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	aaatcgcccg	cagaagcgcg	9240
gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	gtggaaaccg	acgccccagc	9300

actcggtccga	gggcaaagga	ataatcagta	ctgacaataa	aaagattctt	gttttcaaga	9360
acttggtcatt	tgtatagttt	ttttatattg	tagttgttct	attttaatca	aatggttagcg	9420
tgatttatat	tttttttgcg	ctcgacatca	tctgcccaga	tgcgaggtta	agtgcgcaga	9480
aagtaatatc	atgcgtcaat	cgtatgtgaa	tgctggtcgc	tatactgctg	tcgattcgat	9540
actaacgccg	ccatccagtg	tcgacggatc	ctaggtgtac	ataaacttta	taaatgaaat	9600
tcataataga	aacgacacga	aattacaaaa	tggaatatgt	tcatagggtta	gacgaaacta	9660
tatacgcaat	ctacatacat	ttatcaagaa	ggagaaaaag	gaggatagta	aaggaataca	9720
ggtaagcaaa	ttgatactaa	tggctcaacg	tgataaggaa	aaagaattgc	actttaacat	9780
taatattgac	aaggaggagg	gcaccacaca	aaaagttagg	tgtaacagaa	aatcatgaaa	9840
ctacgattcc	taatttgata	ttggaggatt	ttctctaaaa	aaaaaaaaat	acaacaaata	9900
aaaaacactc	aatgacctga	ccatttgatg	gagtttaagt	caataccttc	ttgaaccatt	9960
tcccataatg	gtgaaagttc	cctcaagaat	tttactctgt	cagaaacggc	cttacgacgt	10020
agtcgatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	agccagcccc	10080
gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgctcccg	gcacccgctt	10140
acagacaagc	tgtgaccgtc	tccgggagct	gcattgtgtca	gagggttttca	ccgtcatcac	10200
cgaaacgcgc	gagacgaaag	ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	10260
taataatggt	ttcttaggac	ggatcgcttg	cctgtaactt	acacgcgcct	cgtatctttt	10320
aatgatggaa	taatttgagg	atttactctg	tgtttattta	tttttatggt	ttgtatttgg	10380
attttagaaa	gtaaataaag	aaggtagaag	agttacggaa	tgaagaaaaa	aaaataaaca	10440
aagggtttaa	aaatttcaac	aaaaagcgta	ctttacatat	atatttatta	gacaagaaaa	10500
gcagattaaa	tagatataca	ttcgattaac	gataagtaaa	atgtaaaatc	acaggatttt	10560
cgtgtgtggt	cttctacaca	gacaagatga	aacaattcgg	cattaatacc	tgagagcagg	10620
aagagcaaga	taaaaggtag	tatttggttg	cgatccccct	agagtctttt	acatcttcgg	10680
aaaacaaaaa	ctattttttc	tttaatttct	ttttttactt	tctattttta	atttatatat	10740
ttatattaaa	aaattttaat	tataattatt	tttatagcac	gtgatgaaaa	ggacccaggt	10800
ggcacttttc	ggggaaatgt	gcgcggaacc	cctatttggt	tattttttcta	aatacattca	10860
aatatgtatc	cgtcatgag	acaataaccc	tgataaatgc	ttcaataata	ttgaaaaagg	10920
aagagtatga	gtattcaaca	tttccgtgtc	gcccttattc	ccttttttgc	ggcattttgc	10980
cttctgtttt	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	11040
ggtgcacgag	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	11100
cgccccgaag	aacgtttttc	aatgatgagc	actttttaag	ttctgctatg	tggcgcggtta	11160
ttatcccgtta	ttgacgcggg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	11220
gacttggttg	agtactcacc	agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	11280
gaattatgca	gtgctgccat	aaccatgagt	gataacactg	cggccaactt	acttctgaca	11340
acgatcggag	gaccgaagga	gctaaccgct	ttttttcaca	acatggggga	tcattgtaact	11400
cgccttgatc	gttggggaacc	ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	11460
acgatgcctg	tagcaatggc	aacaacgttg	cgcaactat	taactggcga	actacttact	11520
ctagcttccc	ggcaacaatt	aatagactgg	atggaggcgg	ataaagtgtc	aggaccactt	11580
ctgcgctcgg	cccttcgggc	tggctgggtt	attgctgata	aatctggagc	cgggtgagcgt	11640
gggtctcgcg	gtatcattgc	agcactgggg	ccagatggta	agccctcccg	tatcgtagtt	11700
atctacacga	cgggcagtca	ggcaactatg	gatgaacgaa	atagacagat	cgtgagata	11760
ggtgcctcac	tgattaagca	ttggtaactg	tcagaccaag	tttactcata	tatactttag	11820
attgatttaa	aacttcattt	ttaatttaaa	aggatctagg	tgaagatcct	ttttgataat	11880
ctcatgacca	aaatccctta	acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	11940
aagatcaaaag	gatcttcttg	agatcctttt	tttctgcgcg	taatctgctg	cttgcaaaaca	12000
aaaaaaccac	cgtaccagc	ggtggtttgt	ttgcgggatc	aagagctacc	aactcttttt	12060
ccgaaggtaa	ctggcttcag	cagagcgcag	ataccaaata	ctgtccttct	agtgtagccg	12120
tagttaggcc	accacttcaa	gaactctgta	gcaccgccta	catacctcgc	tctgctaata	12180
ctgttaccag	tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	12240
cgatagttaac	cggataaggc	gcagcggtcg	ggctgaacgg	gggggttcgtg	cacacagccc	12300
agcttgagac	gaacgacctc	caccgaactg	agatacctac	agcgtgagca	ttgagaaagc	12360
gccacgcttc	ccgaaggagg	aaaggcggac	aggtatccgg	taagcggcag	ggtcgggaaca	12420
ggagagcgca	cgagggagct	tccagggggg	aacgcctggt	atcttttatag	tcctgtcggg	12480
tttcgccacc	tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	gccgagccta	12540
tggaaaaaacg	ccagcaacgc	ggccttttta	cggttcctgg	ccttttgctg	gccttttgct	12600
cacatgttct	ttcctgcggt	atccctctgat	tctgtggata	accgtattac	cgccttttgag	12660

tgagctgata	ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	gagcgaggaa	12720
gcggaagagc	gccaataacg	caaaccgcct	ctccccgcgc	gttggccgat	tcattaatgc	12780
agctggcacg	acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	aattaatgtg	12840
agttacctca	ctcattaggc	accccaggct	ttacacttta	tgcttccggc	tcctatgttg	12900
tgtggaattg	tgagcggata	acaatttcac	acaggaaaca	gctatgacca	tgat	12954

<210> 30

<211> 10329

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN608 plasmid for overexpression of DAK1 and GCY1

<400> 30

agtcgtgtct	taccgggttg	gactcaagac	gatagttacc	ggataaggcg	cagcggtcgg	60
gctgaacggg	gggttcgtgc	acacagccca	gcttgagcgc	aacgacctac	accgaactga	120
gataacctaca	gcgtgagcat	tgagaaagcg	ccacgcttcc	cgaagggaga	aaggcggaca	180
ggtatccggg	aagcggcagg	gtcggaaacag	gagagcgcac	gagggagcct	ccagggggga	240
acgcctggta	tctttatagt	cctgtcgggt	ttcgccacct	ctgacttgag	cgtcgatttt	300
tgtgatgctc	gtcagggggg	cagagcctat	ggaaaaacgc	cagcaacgcg	gcctttttac	360
ggttcctggc	cttttgctgg	ccttttgctc	acatgttctt	tcctgcgtta	tcccctgatt	420
ctgtggataa	ccgtattacc	gcctttgagt	gagctgatac	cgctcgccgc	agccgaacga	480
ccgagcgcag	cgagtcagtg	agcgagggaag	cggaagagcg	cccaatacgc	aaaccgcctc	540
tccccgcgcg	ttggccgatt	cattaatgca	gctggcacga	caggtttccc	gactggaaag	600
cgggcagtga	gcgcaacgca	attaatgtga	gttacctcac	tcattaggca	ccccaggcct	660
tacactttat	gcttcgggct	cctatgttgt	gtggaattgt	gagcggataa	caatttcaca	720
caggaaacag	ctatgaccat	gattacgcca	agctcggaat	taaccctcac	taaagggaac	780
aaaagctggg	taccgggccc	cccctcgagt	accggtttct	tcttcagatt	ccctcatgga	840
gaaagtgcgg	cagatgtata	tgacagagtc	gccagtttcc	aagagacttt	attcaggcac	900
ttccatgata	ggcaagagag	aagaccacga	gatgttggtg	tcctagttac	acatgggtatt	960
tattccagag	tattcctgat	gaaatgggtt	agatggacat	acgaagagtt	tgaatcgttt	1020
accaatgttc	ctaacgggag	cgtaatgggt	atggaaactgg	acgaatccat	caatagatac	1080
gtcctgagga	ccgtgctacc	caaatggact	gattgtgagg	gagacctaac	tacatagtgt	1140
ttaaagatta	cggatattta	acttacttag	aataatgccca	tttttttgag	ttataataat	1200
cctacgttag	tgtgagcggg	attttaaactg	tgaggacctt	aatacattca	gacacttctg	1260
cggtatcacc	ctacttattc	ccttcgagat	tatatctagg	aaccatcag	gttgggtggaa	1320
gattacccgt	tctaagaactt	ttcagcttcc	tctattgatg	ttacacctgg	acaccccttt	1380
tctggcatcc	agtttttaat	cttcagtggc	atgtgagatt	ctccgaaatt	aattaaagca	1440
atcacacaaat	tctctcggat	accacctcgg	ttgaaactga	caggtgggtt	gttacgcattg	1500
ctaattgcaaa	ggagcctata	tacctttggc	tcggctgctg	taacagggaa	tataaagggc	1560
agcataattt	aggagtttag	tgaacttgca	acatttacta	ttttcccttc	ttacgtaaat	1620
atttttcttt	ttaattctaa	atcaatcttt	ttcaattttt	tgtttgatt	cttttcttgc	1680
ttaaatctat	aactacaaaa	aacacataca	taaatctaga	aaatgtccgc	taaatcgttt	1740
gaagtcacag	atccagtcaa	ttcaagtctc	aaagggtttg	cccttgctaa	cccctccatt	1800
acgctggtec	ctgaagaaaa	aattctcttc	agaaagaccg	attccgacaa	gatcgcatga	1860
atctctgggtg	gtggtagtg	acatgaacct	acacacgcgc	gtttcattgg	taagggtatg	1920
ttgagtggcg	ccgtgggttg	cgaaattttt	gcacccctt	caacaaaaca	gatttttaaat	1980
gcaatccgtt	tagtcaatga	aaatgcgtct	ggcgttttat	tgattgtgaa	gaactacaca	2040
ggtgatgttt	tgcatthttg	tctgtccgct	gagagagcaa	gagccttggg	tattaactgc	2100
cgcgttgctg	tcataggtga	tgatgttgca	gttggcagag	aaaagggtgg	tatggttggt	2160
agaagagcat	tggcaggtac	cgttttggtt	cataagattg	taggtgcctt	cgcagaagaa	2220
tattctagta	agtatggctt	agacgggtaca	gctaaagtgg	ctaaaattat	caacgacaat	2280
ttggtgacca	ttggatcttc	tttagaccat	tgtaaagtgc	ctggcaggaa	attcgaaagt	2340
gaattaaacg	aaaaacaaat	ggaattgggt	atgggtattc	ataacgaacc	tggtgtgaaa	2400

gttttagacc	ctattccttc	taccgaagac	ttgatctcca	agtatatgct	acaaaaacta	2460
ttggatccaa	acgataagga	tagagctttt	gtaaagtttg	atgaagatga	tgaagttgtc	2520
ttgttagtta	acaatctcgg	cgggtgtttct	aattttgtta	ttagttctat	cacttccaaa	2580
actacggatt	tottaaagga	aaattacaac	ataaccccg	ttcaaacaat	tgctggcaca	2640
ttgatgacct	ccttcaatgg	taatgggttc	agtatcacat	tactaaacgc	cactaaggct	2700
acaaaggctt	tgcaatctga	ttttgaggag	atcaaatacag	tactagactt	gttgaacgca	2760
tttacgaacg	caccgggctg	gccaatgca	gattttgaaa	agacttctgc	cccatctgtt	2820
aacgatgact	tgttacataa	tgaagtaaca	gcaaaggccg	tcggtacct	tgactttgac	2880
aagtttgctg	agtggatgaa	gagtggtgct	gaacaagtta	tcaagagoga	accgcacatt	2940
acggaactag	acaatcaagt	tggtgatggg	gattgtgggt	acactttagt	ggcaggagtt	3000
aaaggcatca	ccgaaaacct	tgacaagctg	tcgaaggact	cattatctca	ggcggttgcc	3060
caaatttcag	atttcattga	aggctcaatg	ggaggactt	ctgggtggtt	atattctatt	3120
cttttgctcg	gtttttcaca	cggattaatt	caggtttgta	aatcaaagga	tgaaccgcgc	3180
actaaggaaa	ttgtggctaa	gtcactcggg	attgcattgg	atactttata	caaataata	3240
aaggcaagga	agggatcctc	caccatgatt	gatgctttag	aaccattcgt	taaagaattt	3300
actgcatcta	aggatttcaa	taaggcggta	aaagctgcag	aggaagggtc	taaatccact	3360
gctacattcg	aggccaaatt	tggcagagct	tcgtatgtcg	gcgattcctc	tcaagtagaa	3420
gatcctgggtg	cagtaggcct	atgtgagttt	ttgaaggggg	ttcaaagcgc	cttgtaagtc	3480
gagacaaatc	gctcttaaat	atatacctaa	agaacattaa	agctatatta	taagcaaaga	3540
tacgtaaatt	ttgcttatat	tattatacac	atatcatatt	tctatatttt	taagatttgg	3600
ttatataatg	tacgtaatgc	aaaggaaata	aattttatac	attattgaac	agcgtccaag	3660
taactacatt	atgtgcacta	atagttttagc	gtcgtgaaga	ctttattgtg	tcgcgaaaag	3720
taaaaatttt	aaaaattaga	gcaccttgaa	cttgcgaaaa	aggttctcat	caactgttta	3780
aaacgtacgt	gtggaagaac	gattacaaca	ggtgttgctc	tctgaggaca	taaaatacac	3840
accgagattc	atcaactcat	tgctggagtt	agcatatcta	caattgggtg	aaatggggag	3900
cgatttgacg	gcatttgctc	ggcatgccgg	tagagggtgtg	gtcaataaga	gcgacctcat	3960
gctataacctg	agaaagcaac	ctgacctaca	ggaaagagtt	actcaagaat	aagaattttc	4020
gttttaaaac	ctaagagtca	ctttaaaatt	tgtatacact	tatttttttt	ataacttatt	4080
taataataaa	aatcataaat	cataagaaat	tcgcgcgctt	acttgaatac	ttcgaaagga	4140
gaccaatttg	gatgtacaac	ccttttttctg	cctttttcct	tcgatatgtt	attgatagct	4200
tcaaagtcct	cagtagacaa	agtaaatatt	ttcctgttcg	ttttgattcg	atcgggattc	4260
acagattttg	gcaagacaac	ataacctctt	tggacgtgcc	agctaataac	aacgtgtccg	4320
ggttgaaacgt	tatttttctt	cgcaatttca	aggataaccg	gttccttcaa	tagtgagaca	4380
tcggtactac	ctaacggaga	ataagcttca	accacaatgc	ctttactttt	acaaaaatta	4440
atcaattcgt	cttgaggtag	taatggatgt	atttcgactt	ggttagcagc	tggcgtaagc	4500
ttattacctt	gagatgctaa	tagatctttc	aggttattta	tagaaaagtt	ggagactcca	4560
acggccttag	ttttaccagt	ctttggtagt	tcctgcatta	attcccaggt	tttgatgaaa	4620
ttccaattgg	tgatatccac	tgcacgagaa	ccatccttct	ttgttggcac	actcaagatg	4680
tcttcatttt	tgatgtaggc	tggatctaata	ctggcaggcc	aatgcaccaa	atataagtct	4740
acgtagtcca	atcctaacct	cttttagtgat	tgatccagcg	ctaacttcagg	ttcgtgggtg	4800
tgtgtacacc	ataactttgt	agtaacaaag	atttcttccc	gaggaacacc	tgaatccttg	4860
atggcttgac	cgacttggtc	ttcattacgg	taaatagcag	cagtatcaat	gtgtcggtag	4920
ccatctttca	aagcgggtta	aacagcctta	taagcatcgt	tctctttcga	ctgccacgta	4980
cctaaacct	tttgagggat	ttgggctcca	gtatttagag	aaaggatttt	cgtagaatca	5040
tgtaaagtag	caggcatttt	ctgcagttta	ttcagtaaat	tttcgatctt	gggaagaaaa	5100
aagcagtaag	cgtgaaaaat	ctaaaagctg	atgtagtaga	agatcctatt	ctttaacaaa	5160
gattgacctt	ttctttttct	tcttggtttg	agtagaaagg	ggaaggaga	atacaagaga	5220
gaggaaaaaa	aggaagataa	aaagagagcg	tgatataaat	gaatatatat	taaacaagag	5280
agattgggaa	ggaaaggatc	aaacaaaccc	aaaaatattt	caaaaaggag	agagagaggg	5340
gagtttggtt	tcaaaacggg	ttattttatt	atgcaagagg	acgtggaaga	aaaagaagaa	5400
ggaagaaaaa	aatttgaaag	aaaaaaacgc	gtggcgggta	aagaagaaaa	tggaaaatag	5460
aggccgggtg	acagagaaat	attgagggtt	aattggaaaa	tatgttaggg	tgaggcatat	5520
gtttttaagg	gttttgagga	tccgataagg	aagaatgtag	gttaaagtgt	gtgcattaat	5580
tgtgtggca	gcttaccgcg	ttccccacac	atttactagt	tctagagcgg	ccgccaccgc	5640
ggtggagctc	caattcgccc	tatagttagt	cgtattacaa	ttcactggcc	gtcgtttttac	5700
aacgtcgtga	ctgggaaaac	cctggcggtta	cccaacttaa	tcgccttgca	gcacatcccc	5760

ccttcgccag	ctggcgtaat	agcgaagagg	cccgcaccga	tcgcccttcc	caacagttgc	5820
gcagcctgaa	tggcgaatgg	cgcgacgcgc	cctgtagecg	cgcattaagc	gcggcggtg	5880
tgggtggttac	gcgcagcgtg	accgctacac	ttgccagcgc	cctagcgccc	gctcctttcg	5940
ctttcttccc	ttcctttctc	gccacgttcg	ccggtagtgt	tagacctgaa	caagggtttac	6000
taaaaatccg	taaagaactt	caattgtacg	ccaacttaag	gcctcgtccc	cgcgggtca	6060
cccggccagc	gacatggagg	cccagaatac	cctccttgac	agtcttgacg	tgcgacagtc	6120
aggggcatga	tgtgactgtc	gcccgtacat	ttagcccata	catcccatg	tataatcatt	6180
tgcattccata	cattttgatg	gccgcacggc	gcgaagcaaa	aattacggct	cctcgtgca	6240
gacctgcgag	cagggaaacg	ctccctcac	agacgcgtga	attgtcccca	cgcgcgccc	6300
ctgtagagaa	atataaaagg	ttaggatttg	ccactgaggt	tcttctttca	tatacttctt	6360
tttaaaatct	tgtaggata	cagttctcac	atcacatccg	aacataaaca	accatgggta	6420
aaaagcctga	actcaccgcg	acgtctgtcg	agaagtttct	gatcgaaaag	ttcgacagcg	6480
tctccgacct	gatgcagctc	tcggagggcg	aagaatctcg	tgctttcagc	ttcgatgtag	6540
gagggcgtgg	atatgtcctg	cgggtaaata	gctgcgcgga	tggtttctac	aaagatcgtt	6600
atgtttatcg	gcactttgca	tcggccgcgc	tcccgattcc	ggaagtgtt	gacattgggg	6660
aattcagcga	gagcctgacc	tattgcatct	cccgcctgct	acaggggtgc	acgttgcaag	6720
acctgcctga	aaccgaactg	cccgtctgtc	tgcagccggt	cgcggaggcc	atggatgcga	6780
tcgctgcggc	cgatcttagc	cagacgagcg	ggttcggccc	attcggaccg	caaggaatcg	6840
gtcaatacac	tacatggcgt	gatttcatat	gcgcgattgc	tgatcccat	gtgtatcact	6900
ggcaaaactgt	gatggacgac	accgtcagtg	cgtccgtcgc	gcaggctctc	gatgagctga	6960
tgctttgggc	cgaggactgc	cccgaagtcc	ggcacctcgt	gcacgcggat	ttcggtcca	7020
acaatgtcct	gacggacaat	ggccgcataa	cagcggtcct	tgactggagc	gaggcgatgt	7080
tcggggatcc	ccaatacgag	gtcgccaaca	tcttcttctg	gaggccgtgg	ttggcttgta	7140
tggagcagca	gacgcgctac	ttcgagcgga	ggcatccgga	gcttgacgga	tcgccgcggc	7200
tccgggcgta	tatgctccgc	attgggtctt	accaactcta	tcagagcttg	gttgacggca	7260
atttcgatga	tgcagcttgg	gcgcagggtc	gatgcgacgc	aatcgctcga	tcgggagccg	7320
ggactgtcgg	gcgtacacaa	atcgcccgcg	gaagcgcggc	cgtctggacc	gatggctgtg	7380
tagaagtact	cgcgatagt	ggaaaccgac	gccccagcac	tcgtccgagg	gcaaaggaat	7440
aatcagtact	gacaataaaa	agattcttgt	tttcaagaac	ttgtcatttg	tatagttttt	7500
ttatatgtga	gttggtctat	tttaatacaa	tgtagcgtg	atttatattt	tttttcgctt	7560
cgacatcatc	tgcccagatg	cgaagttaag	tgcgcagaaa	gtaatatcat	gcgtcaatcg	7620
tatgtgaatg	ctggtcgcta	tactgctgtc	gattcgatac	taacgccgcc	atccagtgtc	7680
gacggatcct	aggtgtacat	aaactttata	aatgaaatcc	ataatagaaa	cgacacgaaa	7740
ttacaaaatg	gaatatgttc	atagggtaga	cgaaactata	tacgcaatct	acatacattt	7800
atcaagaagg	agaaaaagga	ggatagtaaa	ggaatacagg	taagcaaatt	gataactaat	7860
gctcaacgtg	ataaggaaaa	agaattgcac	tttaacatta	atattgacaa	ggaggagggc	7920
accacacaaa	aagttaggtg	taacagaaaa	tcatgaaact	acgattccta	atttgatatt	7980
ggaggatttt	ctctaaaaaa	aaaaaaatac	aacaaataaa	aaacactcaa	tgacctgacc	8040
atttgatgga	gtttaagtea	ataccttctt	gaaccatttc	ccataatggg	gaaagttccc	8100
tcaagaattt	tactctgtca	gaaacggcct	tacgacgtag	tcgatatggg	gcactctcag	8160
tacaatctgc	tctgatgccg	catagttaag	ccagccccga	caccgcgcaa	caccgcgtga	8220
cgcgcctga	cgggcttgtc	tgtctccggc	atccgcttac	agacaagctg	tgaccgtctc	8280
cgggagctgc	atgtgtcaga	ggttttcacc	gtcatcaccg	aaacgcgcga	gacgaaaggg	8340
cctcgtgata	cgcctatttt	tataggttaa	tgtcatgata	ataatggttt	cttaggacgg	8400
atcgcttgcc	tgtaacttac	acgcgcctcg	tatcttttaa	tgatggaata	atttgggaat	8460
ttactctgtg	tttattttatt	tttatgtttt	gtatttggtt	tttagaaagt	aaataaagaa	8520
ggtagaagag	ttacggaatg	aagaaaaaaa	aataaacaaa	ggttttaaaa	atttcaacaa	8580
aaagcgtact	ttacatatat	atttattaga	caagaaaagc	agattaaata	gatatacatt	8640
cgattaacga	taagtaaaat	gtaaaatcac	aggatttttc	tgtgtggtct	tctacacaga	8700
caagatgaaa	caattcggca	ttaataacct	agagcaggaa	gagcaagata	aaaggtagta	8760
tttggttgccg	atccccctag	agtcttttac	atcttcggaa	aacaaaaact	attttttctt	8820
taatttcttt	ttttactttc	tatttttaat	ttatatattt	atattaaaaa	attttaaatta	8880
taattatttt	tatagcacgt	gatgaaaagg	acccaggtgg	cacttttcgg	ggaaatgtgc	8940
gcggaacccc	tatttgttta	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	9000
aataaccctg	ataaatgctt	caataatatt	gaaaaaggaa	gagtatgagt	attcaacatt	9060
tccgtgtcgc	ccttattccc	ttttttgcgg	cattttgcct	tcctgttttt	gctcaccag	9120

aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	tgcacgagt	ggttacatcg	9180
aactggatct	caacagcggg	aagatccttg	agagttttcg	ccccgaagaa	cgttttccaa	9240
tgatgagcac	ttttaaaagt	ctgctatgtg	gcgcgggtatt	atcccgtatt	gacgccgggc	9300
aagagcaact	cggtcgcgcg	atacactatt	ctcagaatga	cttggttgag	tactcaccag	9360
tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	attatgcagt	gctgccataa	9420
ccatgagtga	taacactgcg	gccaacttac	ttctgacaac	gatcggagga	ccgaaggagc	9480
taaccgcttt	ttttcacaac	atgggggata	atgtaactcg	ccttgatcgt	tgggaaccgg	9540
agctgaatga	agccatacca	aacgacgagc	gtgacaccac	gatgcctgta	gcaatggcaa	9600
caacgttgcg	caaactatta	actggcgaac	tacttactct	agcttcccgg	caacaattaa	9660
tagactggat	ggaggcggat	aaagttgcag	gaccacttct	gcgctcggcc	cttcgggctg	9720
gctggtttat	tgctgataaa	tctggagccg	gtgagcgtgg	gtctcgcggg	atcattgcag	9780
cactggggcc	agatggtaag	ccctcccgtg	tcgtagttat	ctacacgacg	ggcagtcagg	9840
caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	tgcctcactg	attaagcatt	9900
ggtaactgtc	agaccaagtt	tactcatata	tacttttagat	tgatttaaaa	cttcattttt	9960
aattttaaag	gatctagggt	aagatccttt	ttgataatct	catgaccaa	atcccttaac	10020
gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	gatcaaagga	tcttcttgag	10080
atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	aaaaccaccg	ctaccagcgg	10140
tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	gaaggtaact	ggcttcagca	10200
gagcgcagat	accaaatact	gtccttctag	tgtagccgta	gttaggccac	cacttcaaga	10260
actctgtagc	accgcctaca	tacctcgctc	tgctaatact	gttaccagtg	gctgctgcca	10320
gtggcgata						10329

<210> 31

<211> 11163

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN616

<400> 31

tttgattaaa	atagaacaac	tacaatataa	aaaaactata	caaatgacaa	gttcttgaaa	60
acaagaatct	ttttattgtc	agtgtgtcag	tctgtctcct	cggccacgaa	gtgcacgcag	120
ttgccggccg	ggtcgcgcag	ggcgaactcc	cgccccacg	gctgtctcgc	gatctcggtc	180
atggccggcc	cggaggcgtc	ccggaagtcc	gtggacacga	cctccgacca	ctcggcgtag	240
agctcgtcca	ggccgcgcac	ccacacccag	gccagggtgt	tgtccggcac	cacctgggtc	300
tggaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	caccggcgaa	gtcgtcctcc	360
acgaagtccc	gggagaaccc	gagccggtcg	gtccagaact	cgaccgctcc	ggcgacgtcg	420
cgcgcgggtg	gcaccggaac	ggcactggtc	aacttgacca	tggttggtta	tgttcggatg	480
tgatgtgaga	actgtatcct	agcaagattt	taaaaggaag	tatatgaaag	aagaacctca	540
gtggcaaatc	ctaacctttt	atatttctct	acaggggcgc	ggcgtgggga	caattcaacg	600
cgtctgtgag	gggagcgttt	ccctgctcgc	aggtctgcag	cgaggagccg	taatttttgc	660
ttcgcgcctg	gcggccatca	aatgttatgg	atgcaaatga	ttatacatgg	ggatgtatgg	720
gctaaatgta	cgggcgacag	tcacatcatg	cccctgagct	gcgcacgtca	agactgtcaa	780
ggaggggtatt	ctgggcctcc	atgtcgtcgg	ccgggtgacc	cggcggggac	gaggccttaa	840
gcggccgcgt	gctagctccg	gattatcgat	gataagctgt	caaacatgag	aattaattcc	900
acggactata	gactatacct	agtatactcc	gtctactgta	cgatacactt	ccgctcaggt	960
ccttgctcct	taacgaggcc	ttaccactct	tttgttactc	tattgatcca	gctcagcaaa	1020
ggcagtgatg	tctaagattc	tatcttcgcg	atgtagtaaa	actagctaga	cggagaaaga	1080
gactagaaat	gcaaaaggca	cttctacaat	ggctgccatc	attattatcc	gatgtgacgc	1140
tgcagcttct	caatgatatt	cgaatacgtc	ttgaggagat	acagcctaata	atccgacaaa	1200
ctgtttttaca	gattttacgat	cgtacttggt	acccatcatt	gaattttgaa	catccgaacc	1260
tgggagtttt	ccctgaaaca	gatagtatat	ttgaacctgt	ataataatat	atagtctagc	1320
gcttttacgga	agacaatgta	tgtatttcgg	ttcctggaga	aactattgca	totattgcat	1380
aggtaatctt	gcacgtcgca	tccccgggtc	attttctgcg	tttccatctt	gcacttcaat	1440

agcatatctt	tggttaacgaa	gcattctgtgc	ttcattttgt	agaacaaaaa	tgcaacgcga	1500
gagcgcta	ttttcaaaca	aagaatctga	gctgcatttt	tacagaacag	aaatgcaacg	1560
cgaagcgct	attttaccaa	cgaagaatct	gtgcttcatt	tttgtaaaac	aaaaatgcaa	1620
cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	attttttacag	aacagaaatg	1680
caacgcgaga	gcgctatttt	accaacaaaag	aatctatact	tctttttttgt	tctacaaaaa	1740
tgcattccga	gagcgctatt	tttctaacaa	agcatcttag	attacttttt	ttctcctttg	1800
tgcgctctat	aatgcagtct	cttgataact	ttttgcactg	taggtccggt	aagggttagaa	1860
gaaggctact	ttggtgtcta	ttttctcttc	cataaaaaaa	gcctgactcc	acttcccgcg	1920
tttactgatt	actagcgaag	ctgcgggtgc	attttttcaa	gataaaggca	tccccgatta	1980
tattctatac	cgatgtggat	tgcgcatact	ttgtgaacag	aaagtgatag	cgttgatgat	2040
tcttcattgg	tcagaaaatt	atgaacgggt	tcttctattt	tgtctctata	tactacgtat	2100
aggaaatggt	tacatttttcg	tattgttttc	gattcactct	atgaatagtt	cttactacaa	2160
tttttttgtc	taaagagtaa	tactagagat	aaacataaaa	aatgtagagg	togagttag	2220
atgcaagttc	aaggagcgaa	aggtggatgg	gtagggtata	tagggatata	gcacagagat	2280
atatagcaaa	gagatacttt	tgagcaatgt	ttgtggaaag	cggtatccgc	aatgccggca	2340
aaagggcgaa	ttgattttga	agagaatgtg	gattttgatg	taattgttgg	gattccattt	2400
ttaataaggc	aataatatta	ggtatgtgga	tatactagaa	gttctcctcg	accgtcgata	2460
tgcggtgtga	aataccgcac	agatgcgtaa	ggagaaaata	ccgcacacag	aaattgtaaa	2520
cgttaaatatt	ttgttaaaat	tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	2580
ataggccgaa	atcggcacaaa	tcccttataa	atcaaaagaa	tagaccgaga	taggggttgag	2640
tggtgttcca	gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	2700
gcgaaaaacc	gtctatcagg	gcgatggccc	actacgtgaa	ccatcacccct	aatcaagttt	2760
tttgggggtcg	aggtgcccga	aagcactaaa	tcggaaccct	aaagggagcc	cccgatttag	2820
agcttgacgg	ggaaagccgg	cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaaggagc	2880
gggcgctagg	gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	caccgcgcgc	2940
gcttaatgcg	ccgctacagg	gcgcgtcgcg	ccattcgcca	ttcaggctgc	gcaactgttg	3000
ggaagggcga	tcgggtgcggg	cctcttcgct	attacgccag	ctggcgaagg	ggggatgtgc	3060
tgcaaggcga	ttaagttggg	taacgccagg	gttttcccag	tcacgacggt	gtaaaacgac	3120
ggccagtga	ttgtaatacg	actcactata	gggcgaattg	gagctccacc	gcggtggcgg	3180
ccgctctaga	actagtctcg	agctcttcaa	ctcaagacgc	acagatatta	taacatctgc	3240
ataataggca	tttgcaagaa	ttactcgtga	gtaaggaaag	agtgaggaa	tatcgcatac	3300
ctgcatttaa	agatgccgat	ttgggcgcga	atcctttatt	ttggcttcac	cctcatacta	3360
ttatcagggc	cagaaaaagg	aagtgtttcc	ctccttcttg	aattgatgtt	accctcataa	3420
agcacgtggc	ctcttatcga	gaaagaaatt	accgtcgctc	gtgatttggt	tgcaaaaaga	3480
acaaaactga	aaaaaaccag	acacgctcga	cttctgtctc	tcctattgat	tgacgcttcc	3540
aatttcgtca	cacaacaagg	tcctagcgac	ggctcacagg	ttttgtaaca	agcaatcgaa	3600
ggttctggaa	tggcgggaaa	gggttttagta	ccacatgcta	tgatgcccac	tgtgatctcc	3660
agagcaaagt	tcgttcgatc	gtactgttac	tctctctctt	tcaaacagaa	ttgtccgaat	3720
cgtgtgacaa	caacagcctg	ttctcacaca	ctcttttctt	ctaaccaagg	gggtggttta	3780
gttttagtaga	acctcgtgaa	acttacattt	acatatatat	aaacttgcat	aaattggtca	3840
atgcaagaaa	tacataattg	gtcttttcta	attcgtagtt	tttcaagttc	ttagatgctt	3900
tctttttctc	tttttttacag	atcatcaagg	aagtaattat	ctacttttta	caacaaatat	3960
atctagactg	cagaaaatgt	ctgaattgaa	cgagaagttg	gctaccgctt	gggaaggttt	4020
caccaagggt	gactggcaaa	acgaagttaa	cgttagagac	ttcatccaaa	agaactacac	4080
cccatacgaa	ggtgacgaat	ctttcttggc	tggtgctacc	gaagctacca	ccaccttggtg	4140
ggacaagggt	atggaagggtg	ttaagttgga	aaacagaacc	cacgctccag	ttgacttcga	4200
caccgctggt	gcttctacca	tcacctctca	cgacgctggg	tacatcaaca	agcaattgga	4260
aaagatcggt	ggttttacaaa	ccgaagctcc	attgaagaga	gctttgatcc	cattcggtgg	4320
tatcaagatg	atcgaagggt	cttgtaaggc	ttacaacaga	gaattggacc	caatgatcaa	4380
gaagattttc	accgaataca	gaaagaccca	caaccaagggt	gttttcgacg	tttactctcc	4440
agacatcttg	agatgtagaa	agtctggtgt	tttgactggt	ttgccagacg	cttacggtag	4500
aggtagaatc	atcgggtgact	acagaagagt	tgctttgtac	ggtatcgact	acttgatgaa	4560
ggacaagttg	gctcaattca	cctctttgca	agctgacttg	gaaaacgggtg	ttactttgga	4620
acaaaccatc	agattgagag	aagaaatcgc	tgaacaacac	agagctttgg	gtcaaatgaa	4680
ggaaatggct	gctaagtacg	gttacgacat	ctctggtcca	gctaccaacg	ctcaagaagc	4740
tatccaatgg	acctacttcg	gttacttggc	tgctgttaag	tctcaaaacg	gtgctgctat	4800

gtcttttcggt	aggacctcta	ccttctttgga	cgtttacatc	gaaagagact	tgaaggctgg	4860
taagatcacc	gaacaagaag	ctcaagaaat	ggttgaccac	ttgggttatga	agttgagaat	4920
ggtttagattc	ttgagaaccc	cagaatacga	cgaattgttc	tctggtgacc	caatctgggc	4980
taccgaatct	atcggtggta	tgggtttgga	cggtagaacc	ttggttacca	agaactcttt	5040
cagattcttg	aacaccttat	acaccatggg	tccatctcca	gaaccaaaca	tgaccatctt	5100
gtggtctgaa	aagttaccat	tgaacttcaa	gaagttcgct	gctaagggtt	ctatcgacac	5160
ctcttctttg	caatacgaaa	acgacgactt	gatgagacca	gacttcaaca	acgacgacta	5220
cgctatcgct	tgttgtgttt	ctccaatgat	cgttggttaag	caaatgcaat	tcttcggtgc	5280
tagagctaac	ttggctaaga	ccatggttga	cgctatcaac	ggtggtgttg	acgaaaagtt	5340
gaagatgcaa	gttggtccaa	agtctgaacc	aatcaagggt	gacgttttga	actacgacga	5400
agttatggaa	agaatggacc	acttcatgga	ctgggttggt	aagcaataca	tcaccgcttt	5460
gaacatcacc	cactacatgc	acgacaagta	ctcttacgaa	gcatacattga	tggcttttgc	5520
cgacagagac	gtaatcagaa	ccatggcttg	tggatatcgct	ggtttgtctg	ttgctgctga	5580
ctctttgtct	gctatcaagt	acgctaagg	taagccaatc	agagacgaag	acggttttggc	5640
tatcgacttc	gaaatcgaag	gtgaataccc	tcaattcggt	aacaacgacc	caagagttga	5700
cgacttggt	gttgacttgg	ttgaaagatt	tatgaagaag	atccaaaagt	tgcacaccta	5760
cagagacgct	atcccaaccc	aatctgtttt	gactatcaca	tctaacgttg	tttacggtta	5820
gaagactggt	aacaccccag	acggtagaag	agctggtgct	ccattcggtc	caggtgctaa	5880
cccaatgcac	ggtagagacc	aaaaggggtgc	tgtagcatct	ttgacctctg	ttgctaagtt	5940
gccattcgct	tacgctaagg	acggtatctc	ttacaccttc	tctatcggtc	caaacgcttt	6000
gggtaaggac	gatgaagtta	gaaagaccaa	cttggtggt	ttgatggacg	gttacttcca	6060
ccaaggaagca	tctatcgaag	gtggtcaaca	cttgaacgta	aatgttatga	acagagaaat	6120
gttggttgac	gctatggaaa	accagaaaa	gtacccacaa	ttgaccatca	gagtttctgg	6180
ttacgctgtt	agattcaact	ctttgaccaa	ggaacaacaa	caagacgtta	tcaccagaac	6240
cttcacccaa	tctatgtaag	tcgagacaaa	tcgctcttaa	atatatacct	aaagaacatt	6300
aaagctatat	tataagcaaa	gatacgtaaa	ttttgcttat	attattatac	acatatcata	6360
ttctatatatt	tttaagattt	ggttatataa	tgtacgtaat	gcaaaggaaa	taaattttat	6420
acattattga	acagcgtcca	agtaactaca	ttatgtgcac	taatagttaa	gcgtcgtgaa	6480
gactttattg	tgtcgcgaaa	agtaaaaatt	ttaaaaatta	gagcaccttg	aacttgcgaa	6540
aaagggtctc	atcaactgtt	taaaacgtac	gtgtggaaga	acgattacaa	caggtgttgt	6600
cctctgagga	cataaaatac	acaccgagat	tcatacaact	attgctggag	ttagcatatc	6660
tacaattggg	tgaaatgggg	agcgatttgc	aggcatttgc	tcggcatgcc	ggtagaggtg	6720
tgggtcaataa	gagcgacctc	atgctatacc	tgagaaagca	acctgacctc	caggaaagag	6780
ttactcaaga	ataagaattt	tcgttttaaa	acctaagagt	cactttaaaa	tttgataaca	6840
cttatttttt	ttataactta	tttaataata	aaaatcataa	atcataagaa	attcgcgcgc	6900
ttagaacata	accttgtgac	cgtattgttc	caagataccc	ttactctttt	ccatgggtttc	6960
cttctttggt	ggcttaacac	cgtccaactt	gtattcttca	cccatagcaa	cccacttgtg	7020
cttaccocaat	tcgtgggtatg	gcaacaattc	gatcttttca	acgttaccca	tgtctctggt	7080
gaactcacc	aatctgtgag	cagagtcgtc	gtcgtcagac	caacctggaa	caacaacgta	7140
tctgatccaa	accttaacgt	tcttggttagc	caagtaacta	gcaaatccca	aggttctgtg	7200
glttagaaaca	ccaaccaagt	tttggtggat	ttcgtcgttc	atttgcttca	agtccaacat	7260
aaccaagtcg	gtaacttcca	acaattcgtc	gataactggg	tcgtatcttc	taacgaaacc	7320
gttggtgtcc	aaacaggtgt	ggataccttc	cttcttacia	gctctgaacc	agtctctaac	7380
aaattcagct	tgcaagatag	cttcaccacc	agaagcggta	acaccacctc	ctgaagcgtt	7440
cataaagtgt	ctgtaggtaa	caacttcctt	catcaagtct	tcaacagtaa	cttccttacc	7500
accgtgggtg	tcccaggtgt	ctctgttgtg	acagtacaaa	catctcatca	aacaaccttg	7560
gaagaagggtg	atgaatctga	tacctggacc	gtcaacagta	ccacaagatt	cgaaagagtg	7620
gattctaccg	ataacagaca	ttttaagctt	ctgcagctta	gattagattg	ctatgctttc	7680
tttctaattga	gcaagaagta	aaaaaagttg	taatagaaca	agaaaaatga	aactgaaact	7740
tgagaaattg	aagaccgttt	attaacttaa	atatcaatgg	gaggtcatcg	aaagagaaaa	7800
aaatcaaaaa	aaaaaatttt	caagaaaaag	aaacgtgata	aaaattttta	ttgccttttt	7860
cgacgaagaa	aaagaaacga	ggcggctctc	ttttcttttt	ccaaaccttt	agtacgggta	7920
attaacgaca	ccctagagga	agaaagaggg	gaaatttagt	atgctgtgct	tgggtgtttt	7980
gaagtggtag	ggcgatgcgc	ggagtcgcag	aaaatctgga	agagtaaaaa	aggagtagaa	8040
acattttgaa	gctatgggtg	gtgggaccgg	tcgagggggg	gcccggtagc	cagcttttgt	8100
tcccttttagt	gagggttaat	tccgagcttg	gcgtaatcat	ggtcatagct	gtttcctgtg	8160

tgaaattggt	atccgctcac	aattccacac	aacataggag	ccggaagcat	aaagtgtaaa	8220
gcctggggtg	cctaattgagt	gaggtaactc	acattaattg	cgttgcgctc	actgcccgtc	8280
ttccagtcgg	gaaacctgtc	gtgccagctg	cattaatgaa	tcggccaacg	cgcggggaga	8340
ggcggtttgc	gtattgggcg	ctcttccgct	tcctcgctca	ctgactcgct	gcgctcggtc	8400
gttcggctgc	ggcgagcggg	atcagctcac	tcaaaggcgg	taatacgggt	atccacagaa	8460
tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaggcc	agcaaaaggc	caggaaccgt	8520
aaaaaggccg	cgttgctggc	gtttttccat	aggctcggcc	cccctgacga	gcatacaaaa	8580
aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	tataaagata	ccaggcggtc	8640
ccccctggaa	gctccctcgt	gcgctctcct	gttcgcgacc	tgcgcgttac	cggataacctg	8700
tccgcctttc	tcccttcggg	aagcgtggcg	ctttctcaat	gctcacgctg	taggtatctc	8760
agttcgggtg	aggtcggtcg	ctccaagctg	ggctgtgtgc	acgaaccccc	cgttcagccc	8820
gaccgctgcg	ccttatccgg	taactatcgt	cttgagtcga	acccggtaag	acacgactta	8880
tcgccactgg	cagcagccac	tggtaacagg	attagcagag	cgaggatgtg	aggcggtgct	8940
acagagttct	tgaagtgggtg	gcctaactac	ggctacacta	gaaggacagt	atttggtatc	9000
tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	gtagctcttg	atccggcaaa	9060
caaaccaccg	ctggtagcgg	tggttttttt	gtttgcaagc	agcagattac	gcgcagaaaa	9120
aaaggatctc	aagaagatcc	tttgatcttt	tctacggggg	ctgacgctca	gtggaacgaa	9180
aactcacggt	aagggatttt	ggtcatgaga	ttatcaaaaa	ggatcttcac	ctagatcctt	9240
ttaaattaaa	aatgaagttt	taaatcaatc	taaagtatat	atgagtaaac	ttggtctgac	9300
agttaccaat	gcttaatcag	tgaggcacct	atctcagcga	tctgtctatt	tcgttcatcc	9360
atagttgcct	gactgcccgt	cgtgtagata	actacgatac	gggagggctt	accatctggc	9420
cccagtgctg	caatgatacc	gcgagaccca	cgctcacccg	ctccagattt	atcagcaata	9480
aaccagccag	ccggaagggc	cgagcgcaga	agtggctcctg	caactttatc	cgcctccatc	9540
cagtctatta	attggtgccg	ggaagctaga	gtaagtagtt	cgccagttaa	tagtttgccg	9600
aacgttgttg	ccattgctac	aggcatcgtg	gtgtcacgct	cgctcgtttg	tatggcttca	9660
ttcagctccg	gttcccaacg	atcaaggcga	gttacatgat	cccccatggt	gtgaaaaaaa	9720
gcggttagct	ccttcgggtc	tccgatcgtt	gtcagaagta	agttggccgc	agtgttatca	9780
ctcatggtta	tggcagcact	gcataattct	cttactgtca	tgccatccgt	aagatgcttt	9840
tctgtgactg	gtgagtactc	aaccaagtca	ttctgagaat	agtgtatgcg	gcgaccgagt	9900
tgctcttgcc	cggcgtcaat	acgggataat	accgcgccac	atagcagAAC	tttaaaagtg	9960
ctcatcattg	gaaaacgttc	ttcggggcga	aaactctcaa	ggatcttacc	gctgttgaga	10020
tccagttcga	tgtAACccac	tcgtgcaccc	aactgatctt	cagcatcttt	tactttcacc	10080
agcgtttctg	ggtgagcaaa	aacaggaagg	caaaatgccg	caaaaaaggg	aataaggggc	10140
acacggaaat	gttgaatact	catactcttc	ctttttcaat	attattgaag	catttatcag	10200
ggttattgtc	tcatgagcgg	atacatattt	gaatgtattt	agaaaaataa	acaaataggg	10260
gttcgcgcga	catttccccg	aaaagtgccg	cctgacgtct	aagaaaccat	tattatcatg	10320
acattaacct	ataaaaatag	gcgtatcacg	aggccctttc	gtctcgcgcg	tttcgggtgat	10380
gacgggtgaaa	acctctgaca	catgcagctc	ccggagacgg	tcacagcttg	tctgtaagcg	10440
gatgccggga	gcagacaagc	ccgtcagggc	gcgtcagcgg	gtgttgggcg	gtgtcggggc	10500
tggetttaact	atgcggcatc	agagcagatt	gtactgagag	tgcaccatat	cgactacgtc	10560
gtaaggccgt	ttctgacaga	gtaaaattct	tgagggaact	ttcaccatta	tgggaaatgg	10620
ttcaagaagg	tattgactta	aactccatca	aatggtcagg	tcattgagtg	ttttttattt	10680
gttgatattt	ttttttttta	gagaaaaatc	tccaatatca	aattaggaat	cgtagtttca	10740
tgattttctg	ttacacctaa	ctttttgtgt	ggtgccctcc	tccttgctca	tattaatgtt	10800
aaagtgcaat	tctttttcct	tatcacgttg	agccattagt	atcaatttgc	ttacctgtat	10860
tcctttacta	tcctcctttt	tctccttctt	gataaatgta	tgtagattgc	gtatatagtt	10920
tcgtctaccc	tatgaacata	ttccattttg	taatttcgtg	tcgtttctat	tatgaatttc	10980
atttataaag	tttatgtaca	cctaggatcc	gtcgacactg	gatggcgggc	ttagtatcga	11040
atcgacagca	gtatagcgac	cagcattcac	atacgattga	cgcatgatat	tactttctgc	11100
gcacttaact	tcgcatctgg	gcagatgatg	tcgaggcgaa	aaaaaatata	aatcacgcta	11160
aca						11163

<210> 32

<211> 7830

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN619

<400> 32

aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	60
cgaaaactct	caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	120
cccaactgat	cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	180
aggcaaaatg	ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	actcatactc	240
ttcctttttc	aatattattg	aagcatttat	caggggttatt	gtctcatgag	cggatacata	300
tttgaatgta	tttagaaaaa	taaacaaata	ggggttcgcg	gcacatttcc	ccgaaaagtg	360
ccacctgaag	totaagaaac	cattattatc	atgacattaa	cctataaaaa	taggcgtatc	420
acgaggccct	ttcgtctcgc	gcgtttcggg	gatgacgggt	aaaacctctg	acacatgcag	480
ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	ggagcagaca	agcccgtcag	540
ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggctggctta	actatgcggc	atcagagcag	600
attgtactga	gagtgcacca	tatcgactac	gtcgtaaggc	cgtttctgac	agagtaaaat	660
tcttgaggga	actttcacca	ttatgggaaa	tggttcaaga	aggtattgac	ttaaactcca	720
tcaaattggc	aggtcattga	gtgtttttta	ttgtttgtat	tttttttttt	ttagagaaaa	780
tcctccaata	tcaaattagg	aatcgtagtt	tcattgattt	ctgttacacc	taactttttg	840
tgtgggtgcc	tcctccttgt	caatattaat	gttaaagtgc	aattcttttt	ccttatcacg	900
ttgagccatt	agtatcaatt	tgtttacctg	tattccttta	ctatcctcct	ttttctcctt	960
cttgataaat	gtatgtagat	tgcgtatata	gtttcgtcta	ccctatgaac	atattccatt	1020
ttgtaatttc	gtgtcgtttc	tattatgaat	ttcatttata	aagtttatgt	acacctagga	1080
tcgctcgaca	ctggatggcg	gcgttagtat	cgaatcgaca	gcagtatagc	gaccagcatt	1140
cacatacgat	tgacgcagta	tattactttc	tgcgcactta	acttcgcata	tgggcagatg	1200
atgtcagagg	gaaaaaaaaa	ataaatcacg	ctaacatttg	attaaaaatg	aacaactaca	1260
atataaaaaa	actatacaaa	tgacaagttc	ttgaaaacaa	gaatcttttt	attgtcagtg	1320
tgtcagtcct	gctcctcggc	cacgaagtgc	acgcagttgc	cggccgggtc	gcgcagggcg	1380
aactcccgc	cccacggctg	ctcgcgcgac	tcggtcattg	ccggcccggg	ggcgtcccgg	1440
aagttcgtgg	acacgacctc	cgaccactcg	gcgtacagct	cgtccaggcc	gcgcacccac	1500
accagggcca	gggtgttgtc	cggcaccacc	tggctcctga	ccgcgctgat	gaacaggggc	1560
acgtcgtccc	ggaccacacc	ggcgaagtcg	tcctccacga	agtcccggga	gaacccgagc	1620
cggtcgggtc	agaactcgac	cgtccggcgg	acgtcgcgcg	cggtgagcac	cggaaacggc	1680
ctggtcaact	tggccatggg	tgttttatgt	cggatgtgat	gtgagaactg	tatcctagca	1740
agatttttaa	aggaagtata	tgaaagaaga	acctcagtg	caaaccctaa	cctttttatat	1800
ttctctacag	gggcgcggcg	tggggacaat	tcaacgcgtc	tgtgagggga	gcgtttccct	1860
gctcgcaggt	ctgcagcgag	gagccgtaat	ttttgcttcg	cgcctgtcgg	ccatcaaaat	1920
gtatggatgc	aaatgattat	acatggggat	gtatgggcta	aatgtacggg	cgacagtcac	1980
atcatgcccc	tgagctgcgc	acgtcaagac	tgtcaaggag	ggtattcttg	gcctccatgt	2040
cgctggccgg	gtgaccggcg	ggggacgagg	ccttaagcgg	ccgcattgta	gctccggatt	2100
atcgatgata	agctgtcaaa	catgagaatt	aattccacgg	actatagact	atacctagta	2160
tactccgtct	actgtacgat	acacttccgc	tcaggctcct	gtcctttaac	gaggccttac	2220
cactcttttg	ttactctatt	gatccagctc	agcaaaggca	gtgtgatcta	agattctatc	2280
ttcgcgatgt	agtaaaacta	gctagaccga	gaaagagact	agaaatgcaa	aaggcacttc	2340
tacaatggct	gccatcatta	ttatccgatg	tgacgctgca	gcttctcaat	gatattcgaa	2400
tacgctttga	ggagatacag	cctaatatcc	gacaaactgt	tttacagatt	tacgatacga	2460
cttggttacc	atcattgaat	tttgaacatc	cgaacctggg	agttttccct	gaaacagata	2520
gtatatttga	acctgtataa	taatatatag	tctagcgctt	tacggaagac	aatgtatgta	2580
tttcggttcc	tggagaaact	attgcatcta	ttgcataggt	aatcttgcac	gtcgcattccc	2640
cggtttcattt	tctgcgtttc	catcttgcac	ttcaatagca	tatctttggt	aacgaagcat	2700
ctgtgcttca	ttttgtagaa	caaaaatgca	acgcgagagc	gctaattttt	caaacaaaga	2760
atctgagctg	cattttttaca	gaacagaaat	gcaacgcgaa	agcgtatatt	taccaacgaa	2820
gaatctgtgc	ttcatttttg	taaaacaaaa	atgcaacgcg	agagcgctaa	tttttcaaac	2880
aaagaatctg	agctgcattt	ttacagaaca	gaaatgcaac	gcgagagcgc	tattttacca	2940

acaaagaatc	tataacttctt	ttttgtttcta	caaaaatgca	tcccagagagc	gctattttttc	3000
taacaaagca	tcttagatta	cttttttttct	cctttgtgcg	ctctataatg	cagtctcttg	3060
ataacttttt	gcactgtagg	tccgttaagg	ttagaagaag	gctactttgg	tgtctatttt	3120
ctcttccata	aaaaaagcct	gactccactt	cccgcgttta	ctgattacta	gcgaagctgc	3180
gggtgcattt	tttcaagata	aaggcatccc	cgattatatt	ctataccgat	gtggattgcg	3240
catactttgt	gaacagaaag	tgatagcggt	gatgattctt	cattgggtcag	aaaattatga	3300
acggtttctt	ctattttgtc	tctatatact	acgtatagga	aatgtttaca	ttttcgtatt	3360
gttttcgatt	cactctatga	atagttctta	ctacaatttt	tttgtctaaa	gagtaatact	3420
agagataaac	ataaaaaatg	tagaggtcga	gttttagatgc	aagttcaagg	agcgaaaggt	3480
ggatgggtag	gttatatagg	gatatagcac	agagatatat	agcaaagaga	tacttttgag	3540
caatgtttgt	ggaaagcggg	attcgcaatg	ccggcaaaag	ggcgaattga	ttttgaagag	3600
aatgtggatt	ttgatgtaat	tgttgggatt	ccatttttaa	taaggcaata	atattaggta	3660
tgtggatata	ctagaagttc	tcctcgaccg	tcgatatgcg	gtgtgaaata	ccgcacagat	3720
gcgtaaggag	aaaataccgc	atcaggaaat	tgtaaacgtt	aatattttgt	taaaattcgc	3780
gttaaatttt	tgttaaatca	gctcattttt	taaccaatag	gccgaaatcg	gcaaaatccc	3840
ttataaatca	aaagaataga	ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	3900
tccactatta	aagaacgtgg	actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	3960
tggcccacta	cgtgaaccat	caccctaate	aagttttttg	gggtcgaggt	gccgtaaage	4020
actaaatcgg	aaccctaaag	ggagcccccg	atttagagct	tgacggggaa	agccggcgaa	4080
cgtggcgaga	aaggaaggga	agaaagcgaa	aggagcgggc	gctagggcgc	tggcaagtgt	4140
agcggtcacg	ctgcgcgtaa	ccaccacacc	cgccgcgctt	aatgcgccgc	tacagggcgc	4200
gtcgcgccat	tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgcgggcctc	4260
ttcgcctatta	cgccagctgg	cgaagggggg	atgtgctgca	aggcgattaa	gttgggtaac	4320
gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattgt	aatacgactc	4380
actatagggc	gaattggagc	tccacgcggg	tggcggccgc	tctagaacta	gtggatcccc	4440
cgggctgcag	gaattcgata	tcaagcttat	cgataccgtc	gacctcgacc	gggtcccacac	4500
accatagctt	caaaatgttt	ctactccttt	tttactcttc	cagattttct	cggactccgc	4560
gcacgcgcgt	accacttcaa	aacacccaag	cacagcatac	taaatttccc	ctctttcttc	4620
ctctagggtg	tcgttaatta	cccgtactaa	aggtttgga	aagaaaaaag	agaccgcctc	4680
gtttcttttt	cttcgtcgaa	aaaggcaata	aaaattttta	tcacgtttct	ttttcttgaa	4740
aatttttttt	tttgattttt	ttctctttcg	atgacctccc	attgatattt	aagttaataa	4800
acggctctca	atttctcaag	tttcagtttc	atttttcttg	ttctattaca	acttttttta	4860
cttcttgctc	attagaaaga	aagcatagca	atctaatact	agctgcagaa	gcttaaaatg	4920
tctgttatcg	gtagaatcca	ctctttcgaa	tcttgttgga	ctgttgacgg	tccagggtatc	4980
agattcatca	ccttcttcca	aggttggttg	atgagatgtt	tgtactgtca	caacagagac	5040
acctgggaca	cccacgggtg	taagggaagt	actgttgaag	acttgatgaa	ggaagtgtgt	5100
acctacagac	actttatgaa	cgcttcagga	ggtggtgtta	ccgcttctgg	tgggtgaagct	5160
atcttgcaag	ctgaatttgt	tagagactgg	ttcagagctt	gtaagaagga	aggtatccac	5220
acctgtttgg	acaccaacgg	tttcgttaga	agatacgacc	cagttatcga	cgaattgttg	5280
gaagttaccg	acttggttat	gttggacttg	aagcaaatga	acgacgaaat	ccaccaaaac	5340
ttggttgggtg	tttctaacca	cagaaccttg	gaatttgcta	agtacttggc	taacaagaac	5400
gttaagggtt	ggatcagata	cgttggttgt	ccagggttgt	ctgacgacga	cgactctgct	5460
cacagattgg	gtgagttcac	cagagacatg	ggtaacgttg	aaaagatcga	attgttgcca	5520
taccacgaat	tgggtaagca	caagtgggtt	gctatgggtg	agaataacaa	gttggacggg	5580
gttaagccac	caaagaagga	aaccatggaa	agagttaagg	gtatcttgga	acaatacggg	5640
cacaaggtta	tgttctaagc	gocgcaattt	cttatgattt	atgattttta	ttattaaata	5700
agttataaaa	aaaataagtg	tatacaaatt	ttaaagtgac	tcttaggttt	taaaacgaaa	5760
attcttattc	ttgagtaact	ctttcctgta	ggtcagggtg	ctttctcagg	tatagcatga	5820
ggtcgctctt	attgaccaca	cctctaccgg	catgccgagc	aaatgcctgc	aaatcgctcc	5880
ccatttcacc	caattgtaga	tatgctaaet	ccagcaatga	gttgatgaat	ctcgggtgtgt	5940
attttatgtc	ctcagaggac	aacacctgtt	gtaatcgttc	ttccacacgt	acccagcttt	6000
tgttcccttt	agtgaggggt	aattccgagc	ttggcgtaat	catggtcata	gctgtttcct	6060
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatag	gagccggaag	cataaagtgt	6120
aaagcctggg	gtgcctaata	agtgaggtaa	ctcacattaa	ttgcgttgcg	ctcactgccc	6180
gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcggccca	acgcgcgggg	6240
agaggcgggt	tgcgtattgg	gcgctcttcc	gcttctcgc	tactgactc	gctgcgctcg	6300

gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatacg	gttatccaca	6360
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccaggaac	6420
cgtaaaaagg	cgcggttgct	ggcgtttttc	cataggctcg	gccccctga	cgagcatcac	6480
aaaaatcgac	gctcaagtca	gaggtggcga	aacccgacag	gactataaag	ataccaggcg	6540
ttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	6600
ctgtccgctt	ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	6660
ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	6720
cccgaaccgt	gcgccttata	cggtaactat	cgtcttgagt	ccaacccggt	aagacacgac	6780
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	6840
gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	6900
atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	6960
aaacaaacca	ccgctggtag	cgggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	7020
aaaaaaggat	ctcaagaaga	tcctttgate	ttttctacgg	ggtctgacgc	tcagtggaac	7080
gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	aaaggatctt	cacctagatc	7140
cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggtct	7200
gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	atttcgttca	7260
tccatagttg	cctgactgcc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	7320
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	7380
ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	ctgcaacttt	atccgcctcc	7440
atccagtcta	ttaattggtg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	7500
cgcaacgttg	ttgccattgc	tacaggcatc	gtggtgtcac	gctcgtcgtt	tggatatggc	7560
tcattcagct	ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgaaaa	7620
aaagcggtta	gctccttcgg	tcctccgata	gttgtcagaa	gtaagttggc	cgcagtgtta	7680
tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcattgccatc	cgtaaagtgc	7740
ttttctgtga	ctggtgagta	ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	7800
agttgctctt	gcccggcgctc	aatacgggat				7830

<210> 33

<211> 9695

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN620

<400> 33

tcgcgcgttt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatacg	ctacgtcgta	aggccgtttc	tgacagagta	aaattcttga	gggaactttc	240
accattatgg	gaaatggttc	aagaaggat	tgacttaaac	tccatcaaat	ggtcagggtca	300
ttgagtgttt	tttatttggt	gtattttttt	tttttttagag	aaaatcctcc	aatatcaaat	360
taggaatcgt	agtttcatga	ttttctgtta	cacctaaact	tttgtgtggt	gccctcctcc	420
ttgtcaatat	taatgttaaa	gtgcaattct	ttttccttat	cacgttgagc	cattagtatc	480
aatttgctta	cctgtattcc	tttactatcc	tcctttttct	ccttcttgat	aaatgtatgt	540
agattgcgta	tatagtttcg	tctaccctat	gaacatatcc	cattttgtaa	tttcgtgtcg	600
tttctattat	gaatttcatt	tataaagttt	atgtacacct	aggatccgtc	gacactggat	660
ggcggcggtta	gtatcgaatc	gacagcagta	tagecaccag	cattcacata	cgattgacgc	720
atgatattac	tttctgcgca	cttaacttcg	catctgggca	gatgatgtcg	aggcgaaaaa	780
aaatataaat	cacgctaaca	tttgattaaa	atagaacaac	tacaatataa	aaaaactata	840
caaatgacaa	gttcttgaaa	acaagaatct	ttttattgtc	agtgtgtcag	tcctgctcct	900
cggccacgaa	gtgcacgcag	ttgccggccg	ggtcgcgcag	ggcgaactcc	cgcgcccacg	960
gctgctcgcc	gatctcggtc	atggccggcc	cggaggcgctc	ccggaagtcc	gtggacacga	1020
cctccgacca	ctcggcgctac	agctcgtcca	ggccgcgcac	ccacacccag	gccagggtgt	1080
tgtccggcac	cacctgggtcc	tggaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	1140

caccggcgaa	gtcgtectcc	acgaagtccc	gggagaaccc	gagccggtcg	gtccagaact	1200
cgaccgctcc	ggcgacgtcg	cgcgcggtga	gcaccggaac	ggcactgggc	aacttggcca	1260
tggttgttta	tgttcggatg	tgatgtgaga	actgtatcct	agcaagattt	taaaaggaag	1320
tatatgaaag	aagaacctca	gtggcaaato	ctaacctttt	atatttctct	acaggggcgc	1380
ggcgtgggga	caattcaacg	cgtctgtgag	gggagcgttt	ccctgctcgc	aggtctgcag	1440
cgaggagccg	taatttttgc	ttcgcgccgt	gcggccatca	aaatgtatgg	atgcaaataa	1500
ttatacatgg	ggatgtatgg	gctaaatgta	cgggcgacag	tcacatcatg	cccctgagct	1560
gcgcacgtca	agactgtcaa	ggaggggtatt	ctgggcctcc	atgtcgtcgg	ccgggtgacc	1620
cggcggggac	gaggccttaa	gcggcgccat	gctagctccg	gattatcgat	gataagctgt	1680
caaacatgag	aattaattcc	acggactata	gactatacct	agtatactcc	gtctactgta	1740
cgatacactt	ccgctcaggt	ccttgtcctt	taacgaggcc	ttaccactct	tttgttactc	1800
tattgatcca	gctcagcaaa	ggcagtgatg	tctaagattc	tatcttcgcg	atgtagtaaa	1860
actagctaga	ccgagaaaga	gactagaaat	gcaaaaggca	cttctacaat	ggctgccatc	1920
attattatoc	gatgtgacgc	tgcagcttct	caatgatatt	cgaatacgtc	ttgaggagat	1980
acagcctaata	atccgacaaa	ctgtttttaca	gattttacgat	cgtacttggt	acccatcatt	2040
gaattttgaa	catccgaacc	tgggagtttt	ccctgaaaca	gatagtatat	ttgaacctgt	2100
ataataatat	atagtctagc	gctttacgga	agacaatgta	tgtatttcgg	ttcctggaga	2160
aactattgca	tctattgcat	aggtaatctt	gcacgtcgca	tccccgggtc	attttctgcg	2220
tttccatctt	gcacttcaat	agcatatctt	tgttaacgaa	gcactctgtc	ttcattttgt	2280
agaacaaaaa	tgcaacgcga	gagcgctaata	ttttcaaaca	aagaatctga	gctgcatttt	2340
tacagaacag	aaatgcaacg	cgaaagcgct	atttttaccaa	cgaagaatct	gtgcttcatt	2400
tttgtaaaac	aaaaatgcaa	cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	2460
attttttacag	aacagaaatg	caacgcgaga	gcgctatttt	accaacaaag	aatctatact	2520
tcttttttgt	tctacaaaaa	tgcacccgga	gagcgctatt	tttctaacaa	agcatcttag	2580
attacttttt	ttctcctttg	tgcgctctat	aatgcagctc	cttgataact	ttttgcaactg	2640
taggtccgtt	aagggttagaa	gaaggctact	ttggtgtcta	ttttctcttc	cataaaaaaa	2700
gctgactcc	acttcccgcg	tttactgatt	actagcgaa	ctgcgggtgc	attttttcaa	2760
gataaaggca	tccccgatta	tattctatac	cgatgtggat	tgcgcatact	ttgtgaacag	2820
aaagtgatag	cgttgatgat	tcttcattgg	tcagaaaatt	atgaacgggt	tcttctattt	2880
tgtctctata	tactacgtat	aggaaatggt	tacatttttcg	tattgttttc	gattcactct	2940
atgaatagtt	cttactacaa	tttttttgtc	taaagagtaa	tactagagat	aaacataaaa	3000
aatgtagagg	tcgagtttag	atgcaagttc	aaggagcgaa	aggtggatgg	gtaggttata	3060
tagggatata	gcacagagat	atatagcaaa	gagatacttt	tgagcaatgt	ttgtggaaaag	3120
cggtatctgc	aatgccggca	aaagggcgaa	ttgattttga	agagaatgtg	gattttgatg	3180
taattgttgg	gattccattt	ttaataaggc	aataatatta	ggtatgtgga	tatactagaa	3240
gttctcctcg	accgtcgata	tgcgggtgtga	aataccgcac	agatgcgtaa	ggagaaaata	3300
ccgcatcagg	aaattgtaaa	cgttaatat	ttgttaaaa	tcgcgttaaa	tttttgttaa	3360
atcagctcat	tttttaacca	atagggcgaa	atcggcacaa	tcccttataa	atcaaaagaa	3420
tagaccgaga	taggggttag	tgttggtcca	gtttggaaca	agagtccact	attaaagaac	3480
gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	gcgatggccc	actacgtgaa	3540
ccatcacccct	aatcaagttt	tttgggggtcg	aggtgcgcta	aagcactaaa	tcggaaccct	3600
aaagggagcc	cccgatthag	agcttgacgg	ggaaagccgg	cgaacgtggc	gagaaaggaa	3660
gggaagaaag	cgaaaggagc	gggcgctagg	gcgctggcaa	gtgtagcggg	cacgctgcgc	3720
gtaaccacca	caccgcgcgc	gcttaatgcg	ccgctacagg	gcgcgtcgcg	ccattcgcca	3780
ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	cctcttcgct	attacgccag	3840
ctggcgaagg	ggggatgtgc	tgcaaggcga	ttaagttggg	taacgccagg	gttttcccag	3900
tcacgacgtt	gtaaaacgac	ggccagtgaa	ttgtaatacg	actcactata	gggcgaattg	3960
gagctccacc	gcggtggcgg	ccgctctaga	actagtggat	cccccgggct	gcagggaattc	4020
gatatcaagc	ttatcgatac	cgtcgacctc	gagctcttca	actcaagacg	cacagatatt	4080
ataacatctg	cataataggc	atttgcaaga	attactcgtg	agtaaggaaa	gagtgaggaa	4140
ctatcgcata	cctgcattta	aagatgccga	tttgggcgcg	aatcctttat	tttggttca	4200
ccctcactact	attatcaggg	ccagaaaaag	gaagtgtttc	cctccttctt	gaattgatgt	4260
taccctcata	aagcacgtgg	cctcttatcg	agaaagaaat	taccgtcgtc	cgtgatttgt	4320
ttgcaaaaag	aacaaaactg	aaaaaaccca	gacacgctcg	acttcctgtc	ttcctattga	4380
ttgcagcttc	caatttcgtc	acacaacaag	gtcctagcga	cggctcacag	gttttgtaac	4440
aagcaatcga	aggttctgga	atggcgggaa	agggtttagt	accacatgct	atgatgcca	4500

ctgtgatctc	cagagcaaag	ttcgttcgat	cgtactgtta	ctctctctct	ttcaaacaga	4560
attgtccgaa	tcgtgtgaca	acaacagcct	gttctcacac	actcttttct	tctaaccaag	4620
ggggtggttt	agtttagtag	aacctcgtga	aacttacatt	tacatatata	taaacttgca	4680
taaattggtc	aatgcaagaa	atacatattt	ggtcttttct	aattcgtagt	ttttcaagtt	4740
cttagatgct	ttctttttct	ctttttttaca	gatcatcaag	gaagtaatta	tctacttttt	4800
acaacaaata	tatctagact	gcagaaaatg	tctgaattga	acgagaagtt	ggctaccgct	4860
tgggaagggt	tcaccaaggg	tgactggcaa	aacgaagtta	acgttagaga	cttcatccaa	4920
aagaactaca	ccccatacga	aggtgacgaa	tctttcttgg	ctgggtgctac	cgaagctacc	4980
accaccttgt	gggacaagggt	tatggaagggt	gttaagttgg	aaaacagAAC	ccacgctcca	5040
gttgacttcg	acaccgctgt	tgcttctacc	atcacctctc	acgacgctgg	ttacatcaac	5100
aagcaattgg	aaaagatcgt	tggttttaca	accgaagctc	cattgaagag	agctttgatc	5160
ccattcgggtg	gtatcaagat	gatcgaagggt	tcttgtaagg	cttacaacag	agaattggac	5220
ccaatgatca	agaagatttt	caccgaatac	agaaagaccc	acaaccaagg	tgttttcgac	5280
gtttacactc	cagacatctt	gagatgtaga	aagtctgggtg	ttttgactgg	tttgccagac	5340
gcttacggta	gaggtagaat	catcgggtgac	tacagaagag	ttgctttgta	cggtatcgac	5400
tacttgatga	aggacaagtt	ggctcaattc	acctctttgc	aagctgactt	ggaaaacggt	5460
gttaacttgg	aacaaaccat	cagattgaga	gaagaaatcg	ctgaacaaca	cagagctttg	5520
ggtcaaataga	aggaaatggc	tgctaagtac	ggttacgaca	tctctgggtc	agctaccaac	5580
gctcaagaag	ctatccaatg	gacctacttc	ggttacttgg	ctgctgttaa	gtctcaaac	5640
ggtgctgcta	tgtctttcgg	taggacctct	accttcttgg	acgtttacat	cgaaagagac	5700
ttgaaggctg	gtaagatcac	cgaacaagaa	gctcaagaaa	tggttgacca	cttggttatg	5760
aagttgagaa	tggttagatt	cttgagaacc	ccagaatacg	acgaattggt	ctctgggtgac	5820
ccaatctggg	ctaccgaatc	tatcgggtgg	atggggttgg	acggtagaac	cttggttacc	5880
aagaactctt	tcagattctt	gaacacctta	tacaccatgg	gtccatctcc	agaaccaaac	5940
atgaccatct	tgtgggtctga	aaagttacca	ttgaacttca	agaagttcgc	tgctaagggt	6000
tctatcgaca	cctcttcttt	gcaatacgaa	aacgacgact	tgatgagacc	agacttcaac	6060
aacgacgact	acgctatcgc	ttgttgtgtt	tctccaatga	tcgttggtaa	gcaaatagcaa	6120
ttcttcgggtg	ctagagctaa	cttggtctaag	accatgttgt	acgctatcaa	cgggtggtgtt	6180
gacgaaaagt	tgaagatgca	agttgggtcca	aagtctgaac	caatcaagggt	tgacgttttg	6240
aactacgacg	aagttatgga	aagaatggac	cacttcatgg	actggttggc	taagcaatac	6300
atcacccgtt	tgaacatcat	ccactacatg	cacgacaagt	actcttacga	agcatcattg	6360
atggctttgc	acgacagaga	cgtaatcaga	accatggctt	gtggatatcgc	tggtttgtct	6420
gttgctgctg	actctttgtc	tgctatcaag	tacgctaagg	ttaagccaat	cagagacgaa	6480
gacggtttgg	ctatcgactt	cgaaatcgaa	ggtgaatacc	ctcaattcgg	taacaacgac	6540
ccaagagttg	acgaacttggc	tggtgacttg	gttgaaagat	ttatgaagaa	gatccaaaag	6600
ttgcacacct	acagagacgc	tatcccaacc	caatctgttt	tgactatcac	atctaacggt	6660
gtttacggta	agaagactgg	taacacccca	gacggtagaa	gagctgggtc	tccattcgggt	6720
ccaggtgcta	acccaatgca	cggtagagac	caaaaggggtg	ctgtagcatc	tttgacctct	6780
gttgctaagt	tgccattcgc	ttacgctaag	gacggtatct	cttacacctt	ctctateggt	6840
ccaaacgctt	tgggtaagga	cgatgaagtt	agaaagacca	acttggctgg	tttgatggac	6900
ggttacttec	accacgaagc	atctatcgaa	ggtgggtcaac	acttgaacgt	aaatgttatg	6960
aacagagaaa	tgttggttga	cgctatggaa	aaccagaaaa	agtaccacaca	attgaccatc	7020
agagtttctg	gttacgctgt	tagattcaac	tctttgacca	aggaacaaca	acaagacgtt	7080
atcaccagaa	ccttcaccca	atctatgtaa	gtcgagacaa	atcgctctta	aatatatacc	7140
taaagaacat	taaagctata	ttataagcaa	agatacgtaa	attttgctta	tattattata	7200
cacatatcat	atltctatat	ttttaagatt	tggttatata	atgtacgtaa	tgcaaaggaa	7260
ataaatttta	tacattattg	aacagcgtcc	aagtaactac	attatgtgca	ctaatagttt	7320
agcgtcgtga	agactttatt	gtgtcgcgaa	aagtaaaaat	tttaaaaatt	agagcacctt	7380
gaacttgcca	aaaaggttct	catcaactgt	ttaaaacgta	cccagctttt	gttcccttta	7440
gtgaggggtta	attccgagct	tggcgtaatc	atgggtcatag	ctgtttcctg	tgtgaaattg	7500
ttatccgctc	acaattccac	acaacatagg	agccggaagc	ataaagtgtta	aagcctgggg	7560
tgcctaataga	gtgaggtaac	tcacattaat	tgcgttgccg	tcactgcccg	ctttccagtc	7620
gggaaacctg	tcgtgccagc	tgcattaatg	aatcggccaa	cgcgcgggga	gaggcgggtt	7680
gcgtattggg	cgtcttcccg	cttctcgcgt	cactgactcg	ctgcgctcgg	tcgttcgggt	7740
gcggcgagcg	gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	7800
taacgcagga	aagaacatgt	gagcaaaaagg	ccagcaaaaag	gccaggaacc	gtaaaaaggc	7860

cgcggttgctg	gcgtttttcc	ataggctcgg	ccccctgac	gagcatcaca	aaaatcgacg	7920
ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	tccccctgg	7980
aagctccctc	gtgcgctctc	ctgttccgac	cctgccgctt	accggatacc	tgtccgcctt	8040
tctcccttcg	ggaagcgtgg	cgctttctca	atgctcacgc	tgtaggtatc	tcagttcggg	8100
gtaggtcggt	cgtccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	8160
cgccttatcc	ggtaactatc	gtcttgagtc	caacccggta	agacacgact	tatcgccact	8220
ggcagcagcc	actggtaaca	ggattagcag	agcgagggtat	gtaggcgggtg	ctacagagtt	8280
cttgaagtgg	tggcctaact	acggctacac	tagaaggaca	gtatttggtg	tctgcgctct	8340
gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	8400
cgtcggtagc	ggtggttttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	8460
tcaagaagat	cctttgatct	tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	8520
ttaagggatt	ttgggtcatga	gattatcaaa	aaggatcttc	acctagatcc	ttttaaatta	8580
aaaatgaagt	tttaaataca	tctaaagtat	atatgagtaa	acttggctctg	acagttacca	8640
atgcttaatc	agtgaggcac	ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	8700
ctgactgccc	gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtgc	8760
tgcaatgata	ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	8820
agccggaagg	gccgagcgca	gaagtgggtcc	tgcaacttta	tccgcctcca	tccagtctat	8880
taattgttgc	cggaagcta	gagtaagtag	ttcgcacggt	aatagtttgc	gcaacgttgt	8940
tgccattgct	acaggcatcg	tgggtgtcacg	ctcgtcggtt	ggtatggctt	cattcagctc	9000
cggttcccaa	cgatcaaggc	gagttacatg	atcccccatg	ttgtgaaaaa	aagcgggttag	9060
ctccttcggg	cctccgatcg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggt	9120
tatggcagca	ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	9180
tgggtgagtac	tcaaccaagt	cattctgaga	atagtgtatg	cggcgaccga	gttgctcttg	9240
cccggcgctca	atacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	9300
tggaaaacgt	tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	9360
gatgtaaccc	actcgtgcac	ccaactgac	ttcagcatct	tttactttca	ccagcgtttc	9420
tgggtgagca	aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	9480
atgttgaata	ctcatactct	tcctttttca	atattattga	agcatttatc	agggttattg	9540
tctcatgagc	ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttcgcgc	9600
cacatttccc	cgaaaagtgc	cacctgacgt	ctaagaaacc	attattatca	tgacattaac	9660
ctataaaaaat	aggcgtatca	cgaggccctt	tcgtc			9695

<210> 34
 <211> 14159
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pRN618

<400>	34	
tggtatacca	acaattcttg	60
gacgactgat	tctattgggt	120
taagaaggta	acatgaggat	180
cgtttacatt	caagccctcc	240
cctgaagtct	aggcccttat	300
ttatatctca	aatttttctt	360
tgaaaaacctt	gcttgagaag	420
tggcggcggt	agtatcgaat	480
catgatatta	ctttctgcgc	540
aaaatataaa	tcacgctaac	600
acaaatgaca	agttcttgaa	660
tgggccacga	agtgcacgca	720
ggctgctcgc	cgatctcggt	780
acctccgacc	actcggcgta	840

ttgtccggca	ccacctgggc	ctggaccgcg	ctgatgaaca	gggtcacgtc	gtcccggacc	900
acaccggcga	agtcgtccct	cacgaagtcc	cgggagaacc	cgagccgggc	ggtccagaac	960
tcgaaccgct	cggcgacgtc	gcgcgcgggt	agcaccggaa	cggcactggg	caacttggcc	1020
atggttgttt	atgttcggat	gtgatgtgag	aactgtatcc	tagcaagatt	ttaaaaggaa	1080
gtatatgaaa	gaagaacctc	agtggcaaat	cctaaccttt	tatatctctc	tacaggggcg	1140
cggcgtgggg	acaattcaac	gcgtctgtga	ggggagcggt	tccttgctcg	caggtctgca	1200
gcgaggagcc	gtaatttttg	cttcgcgcgc	tgcggccatc	aaaatgtatg	gatgcaaatg	1260
attatacatg	gggatgtatg	ggctaaatgt	acgggcgcga	gtcacatcat	gcccctgagc	1320
tgcgcacgtc	aagactgtca	aggagggtat	tctgggcctc	catgtcgcgt	gccgggtgac	1380
ccggcgggga	cgaggcctta	agcggccgcg	tgctagctcc	ggattatcga	tgataagctg	1440
tcaaaccatga	gaattaattc	cacggactat	agactatacc	tagtatactc	cgtctactgt	1500
acgatacaact	tcgcgtcagg	tccttgctct	ttaacgaggg	cttaccactc	ttttgttact	1560
ctattgatecc	agctcagcaa	aggcagtgtg	atctaagatt	ctatcttcgc	gatgtagtaa	1620
aactagctag	acogagaaa	agactagaaa	tgcaaaaggc	acttctacaa	tggctgccat	1680
cattattatc	cgatgtgacg	ctgcagcttc	tcaatgatat	tcgaatacgc	tttgaggaga	1740
tacagcctaa	tatccgacaa	actgttttac	agatttacga	tcgtacttgt	tacccatcat	1800
tgaattttga	acatccgaac	ctgggagttt	tccttgaaac	agatagtata	tttgaacctg	1860
tataataata	tatagtctag	cgctttacgg	aagacaatgt	atgtatttcg	gttcctggag	1920
aaactattgc	atctattgca	taggtaatct	tgcacgtcgc	atccccggtt	cattttctgc	1980
gtttccatct	tgcacttcaa	tagcatatct	ttgttaacga	agcatctgtg	cttcattttg	2040
tagaacaaaa	atgcaacgcg	agagcgctaa	tttttcaaac	aaagaatctg	agctgcattt	2100
ttacagaaca	gaaatgcaac	gcgaaagcgc	tattttacca	acgaagaatc	tgtgcttcat	2160
ttttgtaaaa	caaaaatgca	acgcgagagc	gctaattttt	caaacaaaga	atctgagctg	2220
catttttaca	gaacagaaat	gcaacgcgag	agcgtatatt	taccaacaaa	gaatctatac	2280
ttcttttttg	ttctacaaaa	atgcatcccg	agagcgctat	ttttctaaca	aagcatctta	2340
gattactttt	tttctccttt	gtgcgctcta	taatgcagtc	tcttgataac	tttttgcact	2400
gtaggtccgt	taaggttaga	agaaggctac	tttgggtgtc	attttctctt	ccataaaaaa	2460
agcctgactc	cacttcccgc	gtttactgat	tactagcgaa	gctgcgggtg	cattttttca	2520
agataaaggc	atccccgatt	atattctata	ccgatgtgga	ttgcgcatac	tttgtgaaca	2580
gaaagtgata	gcgttgatga	ttcttcattg	gtcagaaaaa	tatgaacggg	ttcttctatt	2640
ttgtctctat	atactacgta	taggaaatgt	ttacattttc	gtattgtttt	cgattcactc	2700
tatgaatagt	tcttactaca	atttttttgt	ctaaagagta	atactagaga	taaacataaa	2760
aatgtagag	gtcgagttaa	gatgcaagtt	caaggagcga	aagggtggatg	ggtaggttat	2820
atagggatat	agcacagaga	tatatagcaa	agagatactt	ttgagcaatg	tttgtggaaa	2880
gcggtattcg	caatgccggc	aaaagggcga	attgattttg	aagagaatgt	ggattttgat	2940
gtaattgttg	ggattccatt	tttaataagg	caataatatt	aggatatgtg	atatactaga	3000
agttctctc	gaccgtcgat	atgcgggtgt	aaataccgca	cagatgcgta	aggagaaaat	3060
accgcacag	gaaattgtaa	acgttaatat	tttgttaaaa	ttcgcgttaa	atttttgtta	3120
aatcagctca	ttttttaacc	aataggccga	aatcggcaaa	atcccttata	aatcaaaaaga	3180
atagaccgag	atagggttga	gtgttggtcc	agtttggaac	aagagtcac	tattaaagaa	3240
cgtggactcc	aacgtcaaag	ggcgaaaaac	cgtctatcag	ggcgatggcc	cactacgtga	3300
accatcacc	taatcaagtt	ttttggggtc	gaggtgccgt	aaagcactaa	atcggaaccc	3360
taaagggagc	ccccgattta	gagcttgacg	gggaaagccg	gcgaacgtgg	cgagaaagga	3420
agggaagaaa	gcgaaaggag	cgggcgctag	ggcgttgcca	agtgtagcgg	tcacgctgcg	3480
cgtaaccacc	acaccgcgcg	cgcttaatgc	gccgctacag	ggcgcgtcgc	gccattcgcc	3540
attcaggctg	cgcaactgtt	gggaaggggc	atcgggtcgg	gcctcttcgc	tattacgcca	3600
gctggcgaa	gggggatgtg	ctgcaaggcg	attaagttgg	gtaacgccag	ggttttccca	3660
gtcacgacgt	tgtaaaacga	cggccagtga	attgtaatac	gactcactat	agggcgaatt	3720
ggagctccac	cgcggtggcg	gccgctctag	aactagtctc	gagctcttca	actcaagacg	3780
cacagatatt	ataacatctg	cataataggg	atttgcaaga	attactcgtg	agtaaggaaa	3840
gagtgaggaa	ctatcgcata	cctgcattta	aagatgccga	tttgggcgcg	aatcctttat	3900
tttggttca	ccctcatact	attatcaggg	ccagaaaaag	gaagtgtttc	cctccttctt	3960
gaattgatgt	taccctcata	aagcacgtgg	cctcttatcg	agaaagaaat	taccgtcgct	4020
cgtgatttgt	ttgcaaaaag	aacaaaactg	aaaaaaccca	gacacgctcg	acttctgtgc	4080
ttcctattga	ttgcagcttc	caatttcgtc	acacaacaag	gtcctagcga	cggctcacag	4140
gttttgtaac	aagcaatcga	aggttctgga	atggcgggaa	agggtttagt	accacatgct	4200

atgatgceca	ctgtgatctc	cagagcaaaag	ttcgttcgat	cgtactgtta	ctctctctct	4260
ttcaaacaga	attgtccgaa	tcgtgtgaca	acaacagcct	gttctcacac	actcttttct	4320
tctaaccaag	ggggtgggtt	agtttagtag	aacctcgtga	aacttacatt	tacatatata	4380
taaacttgca	taaattggtc	aatgcaagaa	atacatattt	ggtcttttct	aattcgtagt	4440
ttttcaagtt	cttagatgct	ttctttttct	cttttttaca	gatcatcaag	gaagtaatta	4500
tctacttttt	acaacaaata	tatctagact	gcagaaaatg	tctgaattga	acgagaagtt	4560
ggctaccgct	tgggaagggt	tcaccaaggg	tgactggcaa	aacgaagtta	acgttagaga	4620
cttcatccaa	aagaactaca	ccccatacga	aggtgacgaa	tctttcttgg	ctggtgctac	4680
cgaagctacc	accaccttgt	gggacaagggt	tatggaagggt	gttaagttgg	aaaacagAAC	4740
ccacgctcca	gttgacttcg	acaccgctgt	tgcttctacc	atcacctctc	acgacgctgg	4800
ttacatcaac	aagcaattgg	aaaagatcgt	tggtttaca	accgaagctc	cattgaagag	4860
agctttgata	ccattcggtg	gtatcaagat	gatcgaagggt	tcttgtaagg	cttacaacag	4920
agaattggac	ccaatgatca	agaagatttt	caccgaatac	agaaagacc	acaaccaagg	4980
tgttttcgac	gtttacactc	cagacatctt	gagatgtaga	aagtctggtg	ttttgactgg	5040
tttgccagac	gcttacggta	gaggtagaat	catcggtgac	tacagaagag	ttgctttgta	5100
cggatatcgac	tacttgatga	aggacaagtt	ggctcaattc	acctctttgc	aagctgactt	5160
ggaaaacggt	gttaacttgg	aacaaaccat	cagattgaga	gaagaaatcg	ctgaacaaca	5220
cagagctttg	ggtcaaata	aggaaatggc	tgctaagtag	ggttacgaca	tctctggtcc	5280
agctaccaac	gctcaagaag	ctatccaatg	gacctacttc	ggttacttgg	ctgctgttaa	5340
gtctcaaaac	ggtgctgcta	tgtctttcgg	taggacctct	accttcttgg	acgtttacat	5400
cgaaagagac	ttgaaggctg	gtaagatcac	cgaacaagaa	gctcaagaaa	tggttgacca	5460
cttggttatg	aagttgagaa	tggttagatt	cttgagaacc	ccagaatacg	acgaattggt	5520
ctctggtgac	ccaatctggg	ctaccgaatc	tatcggtgggt	atgggtttgg	acggtagAAC	5580
cttggttacc	aagaactctt	tcagattctt	gaacacctta	tacaccatgg	gtccatctcc	5640
agaaccaaac	atgaccatct	tgtgggtctga	aaagttacca	ttgaacttca	agaagtctgc	5700
tgctaagggt	tctatcgaca	cctcttcttt	gcaatacgaa	aacgacgact	tgatgagacc	5760
agacttcaac	aacgacgact	acgctatcgc	ttgttggtgt	tctccaatga	tcgttggtaa	5820
gcaaatacaa	ttcttcgggtg	ctagagctaa	cttggtctaag	accatgttgt	acgctatcaa	5880
cgggtggtgt	gacgaaaagt	tgaagatgca	agttggtcca	aagtctgaac	caatcaaggg	5940
tgacgttttg	aactacgacg	aagttatgga	aagaatggac	cacttcatgg	actggttggc	6000
taagcaatac	atcacgctt	tgaacatcat	ccactacatg	cacgacaagt	actcttacga	6060
agcatcattg	atggctttgc	acgacagaga	cgtaatcaga	accatggctt	gtggtatcgc	6120
tggtttgtct	gttgctgctg	actctttgtc	tgctatcaag	tacgctaagg	ttaagccaat	6180
cagagacgaa	gacggtttgg	ctatcgactt	cgaaatcgaa	ggtgaatacc	ctcaattcgg	6240
taacaacgac	ccaagagttg	acgacttggc	tgttgacttg	gttgaaagat	ttatgaagaa	6300
gatecaaaaag	ttgcacacct	acagagacgc	tatcccaacc	caatctgttt	tgactatcac	6360
atctaacggt	gtttacggta	agaagactgg	taacacccca	gacggtagaa	gagctggtgc	6420
tccattcggt	ccagggtgcta	acccaatgca	cggtagagac	caaaagggtg	ctgtagcatc	6480
tttgacctct	gttgctaagt	tgccattcgc	ttacgctaag	gacggtatct	cttacacctt	6540
ctctatcggt	ccaaacgctt	tgggtaagga	cgatgaagtt	agaaagacca	acttggttgg	6600
tttgatggac	ggttacttcc	accacgaagc	atctatcgaa	ggtggtcaac	acttgaacgt	6660
aatgttatg	aacagagaaa	tgttggttga	cgctatggaa	aaccacagaa	agtacccaca	6720
attgaccatc	agagtttctg	gttacgctgt	tagattcaac	tctttgacca	aggaacaaca	6780
acaagacggt	atcaccagaa	ccttcaccca	atctatgtaa	gtcgagacaa	atcgctctta	6840
aatatatacc	taaagaacat	taaagctata	ttataagcaa	agatacgtaa	attttgctta	6900
tattattata	cacatatcat	atctctatat	ttttaagatt	tggttatata	atgtacgtaa	6960
tgcaaaggaa	ataaatttta	tacattattg	aacagcgtcc	aagtaactac	attatgtgca	7020
ctaatagttt	agcgtcgtga	agactttatt	gtgtcgcgaa	aagtaaaaat	tttaaaaatt	7080
agagcacctt	gaacttgoga	aaaaggttct	catcaactgt	ttaaaacgta	cgtgtggaag	7140
aacgattaca	acagggtgtg	tcctctgagg	acataaaata	cacaccgaga	ttcatcaact	7200
cattgctgga	gttagcatat	ctacaattgg	gtgaaatggg	gagcgatttg	caggcatttg	7260
ctcggcacgc	cggtagagggt	gtggtcaata	agagcgacct	catgctatac	ctgagaaagc	7320
aacctgacct	acaggaaaga	gttactcaag	aataagaatt	ttcgttttta	aacctaaagag	7380
tcactttaaa	atgtgtatac	acttattttt	tttataactt	atttaataat	aaaaatcata	7440
aatcataaga	aattcgcgcg	cttagaacat	aaccttggtga	ccgtattggt	ccaagatacc	7500
cttaactctt	tccatgggtt	ccttcttttg	tggcttaaca	ccgtccaact	tgtattcttc	7560

acccatagca	acccacttgt	gottacccaa	ttcgtggtat	ggcaacaatt	cgatcttttc	7620
aacgttacct	atgtctctgg	tgaactcacc	caatctgtga	gcagagtcgt	cgtcgtcaga	7680
ccaacctgga	acaacaacgt	atctgatcca	aacottaacg	ttcttggttag	ccaagtactt	7740
agcaaattec	aaggttctgt	ggttagaaac	accaaccaag	ttttggtgga	tttcgtcgtt	7800
catttgcttc	aagtccaaca	taaccaagtc	ggtaacttec	aacaattcgt	cgataactgg	7860
gtcgtatctt	ctaacgaaac	cgttggtgtc	caaacagggtg	tggatacctt	ccttcttaca	7920
agctctgaac	cagtctctaa	caaattcagc	ttgcaagata	gcttcaccac	cagaagcggc	7980
aacaccacct	cctgaagcgt	tcataaagtg	tctgtaggta	acaacttcct	tcatacagtc	8040
ttcaacagta	acttccttac	caccgtgggt	gtcccagggtg	tctctgttgt	gacagtacaa	8100
acatctcatc	aaacaacctt	ggaagaagggt	gatgaatctg	atacctggac	cgtcaacagt	8160
accacaagat	tcgaaagagt	ggattctacc	gataacagac	attttaagct	tctgcagctt	8220
agatttagatt	gctatgcttt	ctttctaatg	agcaagaagt	aaaaaaagtt	gtaatagaac	8280
aagaaaaatg	aaactgaaac	ttgagaaatt	gaagaccggt	tattaactta	aatatcaatg	8340
ggaggctcatc	gaaagagaaa	aaaatcaaaa	aaaaaaattt	tcaagaaaaa	gaaacgtgat	8400
aaaaattttt	attgcctttt	tcgacgaaga	aaaagaaacg	aggcggcttc	ttttttcttt	8460
tccaaacctt	tagtacgggt	aattaacgac	accctagagg	aagaaagagg	ggaaatttag	8520
tatgctgtgc	ttgggtgttt	tgaagtggta	cggcgatgcg	cggagtcgga	gaaaatctgg	8580
aagagtaaaa	aaggagttaga	aacattttga	agctatgggtg	tgtgggaccg	gtcgaggggg	8640
ggcccgggtac	ccagcttttg	ttcccttttag	tgagggttaa	ttccgagctt	ggcgtaatca	8700
tggctcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	caacatagga	8760
gccggaagca	taaagtgtaa	agcctgggggt	gcctaattgag	tgaggtaact	cacattaatt	8820
gcgttgccgt	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagct	gcattaatga	8880
atcggccaac	gcgcgggggag	aggcggtttg	cgtattgggc	gctcttcgcg	ttcctcgctc	8940
actgactcgc	tgcgctcgggt	cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	9000
gtaatacgggt	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaaggc	9060
cagcaaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	taggctcggc	9120
ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	ggtggcgaaa	cccgcagga	9180
ctataaagat	accaggcggt	ccccctgga	agctccctcg	tgcgctctcc	tgttccgacc	9240
ctgcgcctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	9300
tgtctacgct	gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	gggctgtgtg	9360
cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtc	9420
aacccggtaa	gacacgaactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	9480
gcgagggtatg	taggcgggtgc	tacagagttc	ttgaagtgggt	ggcctaacta	cggctacact	9540
agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	9600
ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggtttttt	tgtttgcaag	9660
cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	9720
tctgacgctc	agtggaacga	aaactcacgt	taagggtattt	tgggtcatgag	attatcaaaa	9780
aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaaatcaat	ctaaagtata	9840
tatgagtaaa	cttgggtctga	cagttaccaa	tgtttaatca	gtgaggcacc	tatctcagcg	9900
atctgtctat	ttcgttcatc	catagttgcc	tgactgcccg	tcgtgtagat	aactacgata	9960
cgggaggggt	taccatctgg	ccccagtgct	gcaatgatac	cgcgagaccc	acgctcaccg	10020
gctccagatt	tatcagcaat	aaaccagcca	gccggaagggt	ccgagcgcag	aagtggctct	10080
gcaactttat	ccgcctccat	ccagtctatt	aattggttgc	gggaagctag	agtaagtagt	10140
tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	10200
tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	10260
tcccccatgt	tgtgaaaaaa	agcggtttagc	tccttcgggtc	ctccgatcgt	tgtcagaagt	10320
aagttggccg	cagtgttatc	actcatgggt	atggcagcac	tgcataattc	tcttactgtc	10380
atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	10440
tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	tacgggataa	taccgcgcca	10500
catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	10560
aggatcttac	cgtgtttgag	atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	10620
tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	10680
gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaaatac	tcatactctt	cctttttcaa	10740
tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	10800
tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	acctgacgtc	10860
taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgtatcac	gaggcccttt	10920

cgtctcgcgc	gtttcgggtga	tgacgggtgaa	aacctctgac	acatgcagct	cccggagacg	10980
gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	cgcgtcagcg	11040
ggtgttggcg	ggtgtcgggg	ctggcttaac	tatgcggcat	cagagcagat	tgtactgaga	11100
gtgcaccata	tcgactacgt	cgtaaggccg	tttctgacag	agtaaaaattc	ttgaggggaa	11160
tttcaccatt	atgggaaatg	gttcaagaag	gtattgactt	aaactccatc	aaatggtcag	11220
gtcattgagt	gttttttatt	tgttgtattt	tttttttttt	agagaaaatc	ctccaatatc	11280
aaattaggaa	tcgtagtttc	atgattttct	gttacaccta	actttttgtg	tggtgccctc	11340
ctccttgtea	atattaatgt	taaagtgcaa	ttctttttcc	ttatcacgtt	gagccattag	11400
tatcaatttg	cttacctgta	ttcctttact	atcctccttt	ttctccttct	tgataaatgt	11460
atgtagattg	cgtatatagt	ttcgtctacc	ctatgaacat	attccatttt	gtaatttcgt	11520
gtcgtttcta	ttatgaattt	catttataaa	gtttatgtac	gaattctatc	cttttgttgt	11580
ttccgggtgt	acaatatgga	cttcctcttt	tctggcaacc	aaaccatac	atcgggattc	11640
ctataatacc	ttcgttggtc	tccttaacat	gtaggtggcg	gaggggagat	atacaataga	11700
acagatacca	gacaagacat	aatgggctaa	acaagactac	accaattaca	ctgcctcatt	11760
gatggtggta	cataacgaac	taatactgta	gccctagact	tgatagccat	catcatatcg	11820
aagtttcact	accctttttc	catttgccat	ctattgaagt	aataatagge	gcatgcaact	11880
tcttttcttt	ttttttcttt	tctctctccc	ccgttggtgt	ctcaccatat	ccgcaatgac	11940
aaaaaaatga	tggaagacac	taaaggaaaa	aattaacgac	aaagacagca	ccaacagatg	12000
tcgttgttcc	agagctgatg	aggggtatct	cgaagcacac	gaaacttttt	ccttccttca	12060
ttcacgcaca	ctactctcta	atgagcaacg	gtatacggcc	ttccttccag	ttacttgaat	12120
ttgaaataaa	aaaagtttgc	tgtcttgcta	tcaagtataa	atagacctgc	aattattaat	12180
cttttgtttc	ctcgtcattg	ttctcgttcc	ctttcttctt	tgtttctttt	tctgcacaat	12240
atttcaagct	ataccaagca	tacaatcaac	tccagctgca	ttaaaatgag	taatcctcaa	12300
aaagctctaa	acgactttct	gtccagtga	tctgttcata	cacatgatag	ttctaggaaa	12360
caatctaata	agcagtcatc	cgacgaagga	cgtcttctat	cacaaccttc	acatcatcac	12420
tctggtggta	ctaacaacaa	taataacaat	aataataata	ataataacag	taacaacaac	12480
aacaacggca	acgatggggg	aatgatgac	gactatgatt	atgaaatgca	agattataga	12540
ccttctccgc	aaagtgcgcg	gcctactccc	acgtatgttc	cacaatattc	tgtagaaaagt	12600
gggactgctt	tcccgattca	agaggttatt	cctagcgcct	acattaacac	acaagatata	12660
aaccataaag	ataacgggtc	gccgagtga	agcagtaata	gagcattcag	gcctagaggg	12720
cagaccacag	tgtcggccaa	cgtgcttaac	attgaagatt	tttacaacaa	tgcagacgat	12780
gcgcatacca	tcccgaggtc	acattttatc	agaaggagaa	gtaggtcgag	ggctacgagt	12840
aatgctgggc	acagtgccaa	tacaggcgcc	acgaatggca	ggactactgg	tgcccaaact	12900
aatatggaaa	gcaatgaatc	accacgtaac	gtccccatta	tggtgaagcc	aaagacatta	12960
taccagaacc	ctcaaacacc	tacagtcttg	cctccacat	accatccaat	taataaatgg	13020
tcttccgtca	aaaacactta	tttgaaggaa	tttttagccg	agtttatggg	aacaatgggt	13080
atgattattt	tcggtagtgc	tgttgtttgt	caggtcaatg	ttgctgggaa	aatacagcag	13140
gacaatttca	acgtggcttt	ggataacctt	aacgttaccg	ggtcttctgc	agaaacgata	13200
gacgctatga	agagtttaac	atccttggtt	tcatccgttg	cgggcgggtac	ctttgatgat	13260
gtggcattgg	gctgggctgc	tgccgtgggtg	atgggctatt	tctgcgctgg	tggtagtgcc	13320
atctcaggtg	ctcatttgaa	tccgtctatt	acattagcca	atttggtgta	tagaggtttt	13380
cccctgaaga	aagttcctta	ttactttgct	ggacaattga	tcggtgcctt	cacaggcgct	13440
ttgatcttgt	ttatttggtg	caaaagggtg	ttacaagagg	catatagoga	ttggtggatg	13500
aatgaaagtg	ttgcggggaat	gttttgcggt	tttccaaage	cttatctaag	ttcaggacgg	13560
caattttttt	ccgaattttt	atgtggagct	atgttacaag	caggaacatt	tgcgctgacc	13620
gatccttata	cgtgtttgtc	ctctgatgtt	ttcccatgga	tgatgtttat	tttgattttc	13680
attatcaatg	cttccatggc	ttatcagaca	ggtacagcaa	tgaatttggc	tcgtgatctg	13740
ggcccacgtc	ttgcactata	tgcagttgga	tttgatcata	aaatgctttg	ggtgcatcat	13800
catcatttct	tttggtttcc	catggtagge	ccattttattg	gtgcgttaat	ggggggggtg	13860
gtttacgatg	tctgtattta	tcagggtcat	gaatctccag	tcaactggtc	tttaccagtt	13920
tataaggaaa	tgattatgag	agcctgggtt	agaaggcctg	gttggaagaa	gagaaataga	13980
gcaagaagaa	catcggacct	gagtgacttc	tcatacaata	acgatgatga	tgaggaaattt	14040
ggagaaagaa	tggtctttca	aaagacaaa	accaagtcct	ctatttcaga	caacgaaaat	14100
gaagcaggag	aaaagaaagt	gcaattttaa	tctgttcage	gcggcaaaa	aacgttttgg	14159

<210> 35
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer FDHuf

<400> 35
tcgaagactc cgaatgaaaa agacatgcc a g

31

<210> 36
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer FDHur

<400> 36
tccggatacc aagttcattt tcaatacacc cca

33

<210> 37
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer FDHdf

<400> 37
atgcatgcag aatgggttctt atgccac

27

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer FDHdr

<400> 38
gaagacagtt ctgttattaa cgacgagcca

30

<210> 39
<211> 5577
<212> DNA
<213> Artificial Sequence

<220>
<223> pRN621

<400> 39

tctgtgcggt	atltcacacc	gcatacaggt	ggcacttttc	ggggaaatgt	gcgcggaacc	60
cctatttggt	tattttttcta	aatacattca	aatatgtatc	cgctcatgag	acaataaacc	120
tgataaatgc	ttcaataata	gcacgtgagg	agggccacca	tggccaagtt	gaccagtgcc	180
gttccgggtgc	tcaccgcgcg	cgacgtcgcc	ggagcgggtcg	agttctggac	cgaccggctc	240
gggtttctccc	gggacttcgt	ggaggacgac	ttcgccgggtg	tgggtccggga	cgacgtgacc	300
ctgttcatca	gcgcgggtcca	ggaccagggtg	gtgcccggaca	acaccctggc	ctgggtgtgg	360
gtgcgcggcc	tggacgagct	gtacgcgcgag	tggtcggagg	tcgtgtccac	gaacttcogg	420
gacgcctccg	ggccgggcoat	gaccgagatc	ggcgagcagc	cgtggggggcg	ggagttcgcc	480
ctgcgcgacc	cgcccgggcaa	ctgcgtgcac	ttcgtggccg	aggagcagga	ctgacacgtg	540
ctaaaaacttc	atlttttaatt	taaaaggatc	taggtgaaga	tcctttttga	taatctcatg	600
acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccggt	agaaaagatc	660
aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	gctgcttgca	aacaaaaaaa	720
ccacgcgtac	cagcgggtggt	ttgtttgccg	gatcaagagc	taccaactct	ttttccgaag	780
gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgtg	gccgtagtta	840
ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	tcgctctgct	aatcctgtta	900
ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggaactc	aagacgatatg	960
ttaccggata	aggcgcagcg	gtcggggtga	acgggggggtt	cgtgcacaca	gcccagcttg	1020
gagcgaacga	cctacaccca	actgagatac	ctacagcgtg	agctatgaga	aagcgcacag	1080
cttcccgaag	ggagaaaaggc	ggacagggtat	ccggtaagcg	gcagggtcgg	aacaggagag	1140
cgacgcaggg	agcttccagg	gggaaacgcc	tggatatctt	atagtcctgt	cgggtttcgc	1200
cacctctgac	ttgagcgtcg	atltttgtga	tgtcgtcag	gggggcggag	cctatggaaa	1260
aacgcacga	acgcggcctt	tttacgggtc	ctgggctttt	gctggccttt	tgtcacatg	1320
ttctttcctg	cgttateccc	tgattctgtg	gataaccgta	ttaccgcctt	tgagtgagct	1380
gataccgctc	gcgcgagccg	aacgaccgag	cgcagcaggt	cagtgcgaga	ggaagcggaa	1440
gagcgcacca	tacgcaaacc	gcctctcccc	gcgcgttgge	cgattcatta	atgcagctgg	1500
cacgacaggt	ttcccgaactg	gaaagcgggc	agtgcgcga	acgcaattaa	tgtgagttag	1560
ctcaactcatt	aggcacccca	ggcttttacac	tttatgcttc	cggctcgtat	gttgtgtgga	1620
attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	cgccaagcta	1680
tttaggtgac	actatagaat	actcaagcta	tgcatacaagc	ttggtaccga	gctcggatcc	1740
actagtaacg	gccgccagtg	tgtcgggaatt	cgccttcga	agactccgaa	tgaaaaagac	1800
atgccagtaa	taaaactatt	ttgatgttat	gcggaatata	ctattcttgg	attattcact	1860
gttaactaaa	agttggagaa	atcactctgc	actgtcaatc	attgaaaaaa	agaacatata	1920
aaagggcaca	aaattgagtc	tttttttaatg	agttcttgc	gaggaaaagtt	tagttaatat	1980
atcattttacg	taaaatatgc	atattcttgt	attgtgcttt	ttttattcat	tttaagcagg	2040
aacaattttac	aagtattgca	acgctaatac	aatcaaaaata	acagctgaaa	attaatatgt	2100
cgaagggaaa	ggtttttgctg	gttcttttacg	aaggtggtaa	gcatgctgaa	gagcaggaaa	2160
agttattggg	gtgtattgaa	aatgaacttg	gtatccggag	ctcgtacgtt	cgaacttaag	2220
gcctcgtccc	cgccgggtca	cccggccagc	gacatggagg	cccagaatac	cctccttgac	2280
agtcttgacg	tgcgcagctc	aggggcatga	tgtgactgtc	gcccgtaacat	ttagcccata	2340
catcccatg	tataatcatt	tgcatacata	catttttgatg	gcgcacggc	gcgaagcaaa	2400
aattacggct	cctcgtgca	gacctgcgag	cagggaacg	ctccctcac	agacgcgttg	2460
aattgtcccc	acgcgcgcgc	cctgtagaga	aatataaaag	gttaggattt	gccactgagg	2520
ttcttctttc	atatacttcc	ttttaaaatc	ttgctaggat	acagttctca	catcacatcc	2580
gaacataaac	aaccgggtag	cccagaacga	cgcgcggtcg	agatccgtcc	cgccaccgcc	2640
gccgacatgg	cggcgggtctg	cgacatcgtc	aatcactaca	tcgagacgag	cacgggtcaac	2700
ttcgtacgg	agccgcagac	tcgcagggag	tggatcgacg	acctggagcg	cctccaggac	2760
cgctacccct	ggctcgtcgc	cgagggtggag	ggcgtcgtcg	ccggcatcgc	ctacgcgcgc	2820
ccctggaagg	cccgcaacgc	ctacgactgg	accgtcgagt	cgacgggtgta	cgtctcccac	2880
cggcaccagc	ggctcggact	gggtccacc	ctctacaccc	acctgctgaa	gtccatggag	2940
accagggct	tcaagagcgt	ggtcgcgcgc	atcggactgc	ccaacgaccc	gagcgtgcgc	3000
ctgcacgagg	cgctcggata	caccgcgcgc	gggacgctgc	gggcagccgg	ctacaagcac	3060
gggggctggc	acgacgtggg	gttctggcag	cgcgacttct	agctgccggc	cccgccccgc	3120
cccgccggc	ccgtcacaca	gatctaatac	atactgacaa	taaaaagatt	cttgttttca	3180
agaacttgtc	atlttgtatag	tttttttata	ttgtagtgtg	tctatlttaa	tcaaagtgtta	3240
gcgtgattta	tatttttttt	cgcctcgaca	tcactctgcc	agatgcgaag	ttaagtgcgc	3300

```

agaaagtaat atcatgcgtc aatcgtatgt gaatgctggt cgetatactg ctgtcgattc 3360
gataactaac ccgccatcca gtgtcgacgg atcctagggtg tacagggccc aagggcgaat 3420
tctgcagata tccatcacac tggcggccgc tcgagcatgc atgcagaatg gttcttatgc 3480
caccagaget tatggacaga agaaataaga gtgattatga gtatttgtga gcagaagttt 3540
tccggctctcc ttttgttctt gttttggcgt attctccact attcgtccat agcacattta 3600
taccttagct aaatatattt taaagcaaaa ttttcgttat ctcttaaaaa atagaagagc 3660
ggtttattaa tatcaaataa ttgaaactgc tgatatggta gctatataca aaatctgctg 3720
tcaaaatttg gcagtaaacg atcttcacgg tagcggttca aataaagagg aaaagtcttt 3780
ctcccttact gtttttctgg aatttggtc gtcgttaata acagaactgt cttcgaaggg 3840
cgaattctgc agatatccat cacactggcg gccgctcgag catgcatcta gagggcccaa 3900
ttcgccctat agtgagtcgt attacaattc actggccgtc gttttacaac gtcgtgactg 3960
ggaaaaccct ggcgttaccc aacttaatcg ccttgacgca catccccctt tcgccagctg 4020
gcgtaatagc gaagaggccc gcaccgatcg ccttcccaa cagttgcgca gcctatacgt 4080
acggcagttt aagggtttaca cctataaaa agagagccgt tatcgtctgt ttgtggatgt 4140
acagagtgat attattgaca cgcggggcg acggatggtg atccccctgg ccagtgcacg 4200
tctgctgtca gataaagtct cccgtgaact ttaccgggtg gtgcatactg gggatgaaag 4260
ctggcgcatg atgaccaccg atatggccag tgtgccggtc tccgttatcg gggaagaagt 4320
ggctgatctc agccaccgcg aaaatgacat caaaaacgcc attaacctga tgttctgggg 4380
aatataaatg tcaggcatga gattatcaaa aaggatcttc acctagatcc ttttcacgta 4440
gaaagccagt ccgcagaaac ggtgctgacc ccggatgaat gtcagctact gggctatctg 4500
gacaagggaa aacgcaagcg caaagagaaa gcaggtagct tgcagtgggc ttacatggcg 4560
atagctagac tgggcggttt tatggacagc aagcgaaccg gaattgccag ctggggcgcc 4620
ctctggtaag gttgggaagc cctgcaaagt aaactggatg gctttctcgc cgccaaggat 4680
ctgatggcgc aggggatcaa gctctgatca agagacagga tgaggatcgt ttcgcatgat 4740
tgaacaagat ggattgcacg caggttctcc ggccgcttg gtggagaggc tatteggcta 4800
tgactgggca caacagacaa tcggctgctc tgatgccgcc gtgttccggc tgtcagcgca 4860
ggggcgcccg gttctttttg tcaagaccga cctgtccggt gccctgaatg aactgcaaga 4920
cgaggcagcg cggctatcgt ggctggccac gacgggcgtt ccttgccgag ctgtgctcga 4980
cgttgctact gaagcgggaa gggactggct gctattgggc gaagtgccgg ggcaggatct 5040
cctgtcatct caccttctc ctgccgagaa agtatccatc atggctgatg caatgcggcg 5100
gctgcatacg cttgatccgg ctacctgcc attcgaccac caagcgaaac atcgcatcga 5160
gcgagcacgt actcggtgg aagccggtct tgcgatcag gatgatctgg acgaagagca 5220
tcaggggctc gcgccagccg aactgttcgc caggetcaag gcgagcatgc ccgacggcga 5280
ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat atcatgggtg aaaatggccg 5340
cttttctgga ttcategact gtggccggct ggggtgtggc gaccgctatc aggacatagc 5400
gttggetacc cgtgatattg ctgaagagct tggcggcgaa tgggctgacc gcttcctcgt 5460
gctttacggg atcggcgtc ccgattcgca gcgcategcc ttctatcgcc ttcttgacga 5520
gttcttctga attattaacg cttacaattt cctgatgcgg tattttctcc ttacgca 5577

```

<210> 40

<211> 5830

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN622

<400> 40

```

tatttcacac cgcatacagg tggcactttt cggggaaatg tgcgcggaac ccctatttgt 60
ttatttttct aaatacatct aaatatgtat ccgctcatga gacaataacc ctgataaatg 120
cttcaataat agcacgtgag gagggccacc atggccaagt tgaccagtgc cgttccgggtg 180
ctcaccgcgc gcgacgtcgc cggagcggtc gagttctgga ccgaccggct cgggttctcc 240
cgggacttcg tggaggacga cttcgccggt gtggtccggg acgacgtgac cctgttcttc 300
agcgcgggtc aggaccaggt ggtgccggac aacaccctgg cctgggtgtg ggtgcgcggc 360
ctggacgagc tgtacgccga gtggtcggag gtcgtgtcca cgaacttccg ggacgcctcc 420

```

gggceggcca	tgaccgagat	cggcgagcag	ccgtgggggc	gggagttcgc	cctgcgcgac	480
ccggccggca	actgcgtgca	cttcgtggcc	gaggagcagg	actgacacgt	gctaaaactt	540
catttttaat	ttaaaaggat	ctaggtgaag	atcctttttg	ataatctcat	gaccaaaatc	600
ccttaacgtg	agtttttcgtt	ccactgagcg	tcagaccccg	tagaaaagat	caaaggatct	660
tcttgagatc	ctttttttct	gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	720
ccagcgggtg	tttgtttgcc	ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	780
ttcagcagag	cgcagatacc	aaatactgtc	cttctagtgt	agccgtagtt	aggccaccac	840
ttcaagaact	ctgtagcacc	gcctacatac	ctcgcctctgc	taatcctgtt	accagtggct	900
gctgccagtg	gcgataagtc	gtgtctttacc	gggttggaact	caagacgata	gttaccggat	960
aaggcgcagc	ggtcgggctg	aacggggggg	tcgtgcacac	agcccagctt	ggagcgaacg	1020
acctacaccg	aactgagata	cctacagcgt	gagctatgag	aaagcgcac	gcttcccgaa	1080
gggagaaagg	cggacaggta	tccggtaagc	ggcagggctc	gaacaggaga	gcgcacgagg	1140
gagcttccag	ggggaaacgc	ctgggtatctt	tatagtcctg	tcgggtttcg	ccacctctga	1200
cttgagcgtc	gattttttgtg	atgctcgtca	ggggggcgga	gcctatggaa	aaacgccagc	1260
aacgcggcct	ttttacgggt	cctgggcttt	tgctggcctt	ttgctcacat	gttctttcct	1320
gcgttatccc	ctgattctgt	ggataaccgt	attaccgcct	ttgagtgagc	tgataccgct	1380
cgcgcagcc	gaacgaccga	gcgcagcgag	tcagtgagcg	aggaagcgga	agagcgccca	1440
atacgcaaac	cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	gcacgacagg	1500
tttcccgaact	ggaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	gtcactcat	1560
taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tgttgtgtgg	aattgtgagc	1620
ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgccaaagt	atttaggtga	1680
cactatagaa	tactcaagct	atgcatcaag	cttgggtaccg	agctcggatc	cactagtaac	1740
ggccgccagt	gtgctggaat	tgcctcttcc	gaagactccg	aatgaaaaag	acatgccagt	1800
aataaaaaata	attgatgtta	tgcggaatat	actattcttg	gattattcac	tgttaactaa	1860
aagttggaga	aatcactctg	cactgtcaat	cattgaaaaa	agaacatat	aaaagggcac	1920
aaaatcgagt	cttttttaat	gagttcttgc	tgaggaaaat	ttagttaata	tatcatttac	1980
ataaaacatg	catattattg	tgltgtactt	tctttattca	ttttaagcag	gaataattac	2040
aagtattgca	acgctaatac	aatcgaaata	acagctgaaa	attaatatgt	cgaagggaaa	2100
ggttttgctg	gttctttatg	aagggtgtaa	gcatgctgaa	gagcaggaaa	agttattggg	2160
gtgtattgaa	aatgaacttg	gtatccggag	ctcgtacgtt	cgaacttaag	gcctcgtccc	2220
cgcgggtca	cccgccagc	gacatggagg	cccagaatac	cctccttgac	agtcttgacg	2280
tgcgcagctc	aggggcatga	tgtgactgtc	gcccgtaacat	ttagcccata	catccccatg	2340
tataatcatt	tgcattccata	cattttgatg	gccgcacggc	gcgaagcaaa	aattacggct	2400
cctcgtctga	gacctgcgag	cagggaacg	ctcccctcac	agacgcgttg	aattgtcccc	2460
acgcgcgcgc	cctgtagaga	aatataaaaag	gttaggattt	gccactgagg	ttcttctttc	2520
atatacttcc	ttttaaaatc	ttgctaggat	acagttctca	catcacatcc	gaacataaac	2580
aacctatgggt	aaggaaaaga	ctcacgtttc	gaggccgcga	ttaaatcca	acatggatgc	2640
tgatttatat	gggtataaat	gggctcgcga	taatgtcggg	caatcagggtg	cgacaatcta	2700
tcgattgtat	gggaagcccg	atgcgccaga	gttgtttctg	aaacatggca	aaggtagcgt	2760
tgccaatgat	gttacagatg	agatggtcag	actaaaactgg	ctgacggaat	ttatgcctct	2820
tccgaccatc	aagcatttta	tccgtactcc	tgatgatgca	tggttactca	ccactgcgat	2880
ccccggcaaa	acagcattcc	aggtattaga	agaatatcct	gattcagggtg	aaaatattgt	2940
tgatgcgctg	gcagtgttcc	tgcgcgggtt	gcattcgatt	cctgtttgta	attgtccttt	3000
taacagcgat	cgcgtatttc	gtctcgtcga	ggcgcaatca	cgaatgaata	acggtttggt	3060
tgatgcgagt	gattttgatg	acgagcgtaa	tggctggcct	gttgaacaag	tctggaaaga	3120
aatgcataag	cttttgccat	tctcacccgga	ttcagtcgtc	actcatgggtg	atctctcact	3180
tgataacctt	atttttgacg	aggggaaatt	aataggttgt	attgatgttg	gacgagtcgg	3240
aatcgcagac	cgataccagg	atcttgccat	cctatggaac	tgctcgggtg	agttttctcc	3300
ttcattacag	aaacggcttt	ttcaaaaata	tgggtattgat	aatcctgata	tgaataaatt	3360
gcagtttcat	ttgatgctcg	atgagttttt	ctaatacagta	ctgacaataa	aaagattctt	3420
gttttcaaga	acttgtcatt	tgtatagttt	ttttatattg	tagttgttct	attttaataca	3480
aatgttagcg	tgatttatat	tttttttcgc	ctcgacatca	tctgcccgaga	tgcgaagtta	3540
agtgcgcaga	aagtaatatc	atgcgtcaat	cgtatgtgaa	tgctggtcgc	tatactgctg	3600
tcgattcgat	actaacgcgc	ccatccagtg	tcgacggatc	ctaggtgtac	agggcccaaa	3660
agggcgaatt	ctgcagatat	ccatcacact	ggcggccgct	cgagcatgca	gaatggttct	3720
tatgccacca	gagcttatgg	acagaagaaa	taagagtgat	tatgagtatt	tgtgagcaga	3780

```

agttttccgg tctccttttg ttcttgtttt ggctgattct ccactattcg tccatagcac 3840
at ttataacct tagctaaata ttttgtaaag caaaattttc gttatctctt aaaaaataga 3900
agagcggttt attaatatca aataattgaa actgctgata tggtagctat atacaaaatc 3960
tgctgtcaaa atttggcagt aaacgatctt cacggtagcg gttcaaataa agaggaaaag 4020
tccttctccc ttactgtttt tctggaattt ggctcgctcg taataacaga actgtcttca 4080
agggcggaatt ctgcagatat ccatacact ggcgccgct cgagcatgca tctagagggc 4140
ccaattcgcc ctatagttag tegtattaca attcactggc cgtcgtttta caacgtcggtg 4200
actgggaaaa ccctggcggt acccaactta atcgccctgc agcacatccc cctttcgcca 4260
gctggcgtaa tagcgaagag gccgcacccg atcgcccttc ccaacagttg cgcagcctat 4320
acgtacggca gtttaagggt tacacctata aaagagagag ccgttatcgt ctgtttgtgg 4380
atgtacagag tgatattatt gacacgcccg ggcgacggat ggtgatcccc ctggccagtg 4440
cacgtctgct gtcagataaa gtctcccgtg aactttaccc ggtgggtgcat atcggggatg 4500
aaagctggcg catgatgacc accgatattg ccagtgtgcc ggtctccgtt atcggggaag 4560
aagtggctga tctcagccac cgcgaaaatg acatcaaaaa cgcattaac ctgatgttct 4620
ggggaatata aatgtcaggc atgagattat caaaaaggat cttcacctag atccttttca 4680
cgtagaaaagc cagtccgcag aaacgggtgt gaccccggtt gaatgtcagc tactgggcta 4740
tctggacaag ggaaaacgca agcgcaaaga gaaagcaggt agcttgcatg gggcttacat 4800
ggcgatagct agactgggcg gttttatgga cagcaagcga accggaattg ccagctgggg 4860
cgccctctgg taagggtggg aagccctgca aagtaaactg gatggctttc tcgccgcaa 4920
ggatctgatg gcgcagggga tcaagctctg atcaagagac aggatgagga tcgtttcgca 4980
tgattgaaca agatggattg cacgcaggtt ctccggccgc ttgggtggag aggtattcg 5040
gctatgactg ggcacaacag acaatcggtt gctctgatgc cgcctgttcc cggctgtcag 5100
cgcagggggc ccgggttctt tttgtcaaga ccgacctgtc cggtgccctg aatgaactgc 5160
aagacgaggg agcgcggcta tcgtggctgg ccacgacggg cgttccttgc gcagctgtgc 5220
tcgacgttgt cactgaagcg ggaagggaact ggetgctatt gggcgaagtg ccggggcagg 5280
atctcctgtc atctcacctt gctcctgccg agaaagtatc catcatggct gatgcaatgc 5340
ggcggtgca tacgcttgat ccggctacct gccattcga ccaccaagcg aaacatcgca 5400
tcgagcgagc acgtactcgg atggaagccg gtcttgctga tcaggatgat ctggacgaag 5460
agcatcaggg gctcgcgcca gccgaactgt tcgccaggt caaggcgagc atgcccgacg 5520
gcgaggatct cgtcgtgacc catggcgatg cctgcttgcc gaatatcatg gtggaaaatg 5580
gccgcttttc tggattcctc gactgtggcc ggctgggtgt ggccgaccgc tatcaggaca 5640
tagcgttggc taaccgtgat attgctgaag agcttggcgg cgaatgggtt gaccgcttcc 5700
tcgtgcttta cggtatcgcc gctcccgtt cgcagcgcct cgccttctat cgccttcttg 5760
acgagttctt ctgaattatt aacgcttaca atttctgat gcggtatttt ctcttacgc 5820
atctgtgcgg 5830

```

<210> 41

<211> 7690

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN558

<400> 41

```

tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcggttg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accataattc cgttttaaga gcttgggtgag cgctaggagt cactgccagg tatcgtttga 240
acacggcatt agtcaggga gtcataacac agtcctttcc cgcaattttc tttttctatt 300
actcttggcc tcctctagta cactctatat ttttttatgc ctcggtaatg attttcattt 360
ttttttttcc acctagcgga tgactctttt tttttcttag cgattggcat tatcacataa 420
tgaattatac attatataaa gtaatgtgat ttcttcgaag aatatactaa aaaatgagca 480
ggcaagataa acgaaggcaa agatgacaga gcagaaagcc ctagtaaagc gtattacaaa 540
tgaaaccaag attcagattg cgatctcttt aaagggtggg cccttagcga tagagcactc 600

```

gatcttccca	gaaaaagagg	cagaagcagt	agcagaacag	gccacacaat	cgcaagtgat	660
taacgtccac	acaggtatag	ggtttctgga	ccatatgata	catgctctgg	ccaagcattc	720
cggtcggtcg	ctaatcgttg	agtgcattgg	tgacttacac	atagacgacc	atcacaccac	780
tgaagaactgc	gggattgctc	tcggtcgaagc	ttttaaaagag	gccctactgg	cgcgtaggag	840
aaaaaggttt	ggatcaggat	ttgcgccttt	ggatgaggca	ctttccagag	cggtggtaga	900
tctttcgaac	aggccgtacg	cagttgtcga	acttgggtttg	caaagggaga	aagtaggaga	960
tctctcttgc	gagatgatcc	cgcattttct	tgaaagcttt	gcagaggcta	gcagaattac	1020
cctccacggt	gattgtctgc	gaggcaagaa	tgatcatcac	cgtagtgaga	gtgcgttcaa	1080
ggctcttgcg	gttgccataa	gagaagccac	ctcgcccaat	ggtaccaacg	atgttccctc	1140
caccaaaggt	gttcttatgt	agtgcacccg	attattttaa	gctgcagcat	acgatatata	1200
tacatgtgta	tatatgtata	cctatgaatg	tcagtaagta	tgtatacgaa	cagtatgata	1260
ctgaagatga	caaggtaatg	catcattcta	tacgtgtcat	tctgaacgag	gcgcgctttc	1320
cttttttctt	tttgcctttt	cttttttttt	ctcttgaact	cgacggatca	tatgcggtgt	1380
gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatac	ggaaattgta	aacgttaata	1440
ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	attttttaac	caataggccg	1500
aaatcggcaa	aatcccttat	aatcaaaaag	aatagaccga	gatagggttg	agtgttggtc	1560
cagtttgtaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	1620
ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatacagt	tttttggggg	1680
cgaggtgccg	taaagcacta	aatcgggaacc	ctaaagggag	cccccgattt	agagcttgac	1740
ggggaaaagcc	ggcattgcga	ataccgcttc	cacaaacatt	gctcaaaaagt	atctctttgc	1800
tatatatctc	tgtgctatat	ccctatatata	cctacccatc	cacctttcgc	tccttgaact	1860
tgcatctaaa	ctcgacctct	acatttttta	tgtttatctc	tagtattact	ctttagacaa	1920
aaaaattgta	gtaagaacta	ttcatagagt	gaatcgaaaa	caatacgaaa	atgtaaacaat	1980
ttcctatacg	tagtatatag	agacaaaata	gaagaaaccg	ttcataattt	tctgaccaat	2040
gaagaatcat	caacgctatc	actttctggt	cacaaagtat	gcgcaatcca	catcggtata	2100
gaatataatc	ggggatgcct	ttatcttgaa	aaaatgcacc	cgcagcttcg	ctagtaatca	2160
gtaaacgcgg	gaagtggagt	caggcttttt	ttatggaaga	gaaaatagac	accaaagtag	2220
ccttcttcta	accttaacgg	acctacagtg	caaaaagtta	tcaagagact	gcattataga	2280
gcgcacaaaag	gagaaaaaaa	gtaatctaag	atgctttggt	agaaaaatag	cgctctcggg	2340
atgcattttt	gtagaacaaa	aaagaagtat	agattctttg	ttggtaaaat	agcgctctcg	2400
cgttgcattt	ctgttctgta	aaaatgcagc	tcagattctt	tgtttgaaaa	attagcgctc	2460
tcgtcgcgtt	gcatttttgt	tttacaaaaa	tgaaacacag	attcttcgtt	ggtaaaaatag	2520
cgctttcgcg	ttgcatttct	gttctgtaaa	aatgcagctc	agattctttg	tttgaaaaat	2580
tagcgtctct	gcgttgcat	tttgttctac	aaaatgaagc	acagatgctt	cgtaaacaaa	2640
gatatgctat	tgaagtgcaa	gatggaaacg	cagaaaatga	accggggatg	cgacgtgcaa	2700
gattacctat	gcaatagatg	caatagtttc	tcaggaaccc	gaaatacata	cattgtcttc	2760
cgtaaagcgc	tagactatat	attattatac	aggttcaaata	atactatctg	tttcaggga	2820
aactcccagg	ttcggatggt	caaaattcaa	tgatgggtaa	caagtacgat	cgtaaatctg	2880
taaaacagtt	tgtcggatat	taggctgtat	ctcctcaaag	cgtattcgaa	tatcattgag	2940
aagctgcagc	gtcacatcgg	ataataatga	tggcagccat	tgtagaagtg	ccttttgcac	3000
ttctagtctc	tttctcggtc	tagctagttt	tactacatcg	cgaagataga	atcttagatc	3060
acactgcctt	tgctgagctg	gatcaataga	gtaacaaaag	agtggtaagg	cctcgttaaa	3120
ggacaaggac	ctgagcggaa	gtgtatcgta	cagtagacgg	agtatctagt	atagtctata	3180
gtccgtggaa	ttaattctca	tctttgacag	cttatcatcg	ataatccgga	gctagcatgc	3240
ggccgccagt	gtgatggata	tctgcagaat	tcgccctttt	aagcttcgta	cgtgtggaag	3300
aacgattaca	acaggtgttg	tcctctgagg	acataaaata	cacaccgaga	ttcatcaact	3360
cattgctgga	gttagcatat	ctacaattgg	gtgaaatggg	gagcgatttg	caggcatttg	3420
ctcggcatgc	cggtagaggt	gtggtcaata	agagcgacct	catgctatac	ctgagaaagc	3480
aacctgacct	acaggaaaga	gttactcaag	aataagaatt	ttcgttttta	aacctaagag	3540
tcacttttaa	atltgtatac	acttattttt	tttataactt	atttaataat	aaaaatcata	3600
aatcataaga	aattcgcgcg	cttaagcagc	ttcaccagcc	tttctagcca	aagattgagc	3660
catcttttca	gcggtagcca	aagcagaaga	ggtcataatg	tccaagtta	cagcgtaagc	3720
tggaagtag	tgagcagcac	cttcaacttc	caaccaaaca	gcggtcttca	aaccagagaa	3780
ttgaccaaca	cctggcaagt	taactggctt	gtcttggtgg	ataacttcga	attgaaactc	3840
ttgcttcaat	ctgtaacctg	gaacgtatgc	ttgaacagct	tcagccattt	cgttgattga	3900
agcttcgatg	tcgtcttgag	atgcttcgct	agacaaaacg	taaacggtgt	ctctcatcat	3960

caatggtggt	tcagctgggt	tcaaaaacgat	gatagcotta	cccttagcag	caccaccaac	4020
aacttcgata	gccctagagg	tggtttcggt	gaattcgctc	atgttagctc	tggtagctgg	4080
accagcagac	ttagaagcga	tagaagcgat	gatttcagcg	tagtgaactc	tagcaactct	4140
tgaaacagca	gcaaccattg	ggatggtagc	ttgaccacca	caggtaacca	tgttaacggt	4200
taattggtca	acgttagctt	ccaagttaac	aactggaaca	cagtatggac	cgatagcagc	4260
tggggtcaag	tcgatcaatc	tgatgtctgg	cttagcttct	ctcaaagcag	cgtcgttctt	4320
aacgtgagca	ccagcagagg	tagcgtcgaa	aacgatgtcg	atgtcagcga	attctggcat	4380
gttcatcaaa	ccgataaac	cttcgtgggt	ggtagcaaca	cccattcttc	tagctctagc	4440
caaaccgtca	gattgtgggt	cgataccaac	cataacagcc	atttccaagt	gttgaccgtg	4500
tcttaggata	ttgatcatca	agtcagtacc	gatgttacca	gaaccgatga	tagcaacctt	4560
tctcttagac	atctgcagtc	tagatatatt	tgttgtaaaa	agtagataat	tacttccttg	4620
atgatctgta	aaaaagagaa	aaagaaagca	tctaagaact	tgaaaaacta	cgaattagaa	4680
aagaccaa	atgtatttct	tgcattgacc	aatttatgca	agtttatata	tatgtaa	4740
taagtttcac	gaggttctac	taaactaaac	caccccttg	gttagaagaa	aagagtgtgt	4800
gagaacaggc	tgttggtg	acacgattcg	gacaattctg	tttgaagag	agagagtaac	4860
agtagcatcg	aacgaacttt	gctctggaga	tcacagtggg	catcatagca	tgtggtacta	4920
aaccctttcc	cgccattcca	gaaccttcga	ttgcttggtt	caaaacctgt	gagccgtcgc	4980
taggaccttg	ttgtgtgacg	aaattggaag	ctgcaatcaa	taggaagaca	ggaagtcgag	5040
cgtgtctggg	ttttttcagt	tttgttcttt	ttgcaaacia	atcacgagcg	acggtaattt	5100
ctttctcgat	aagaggccac	gtgctttatg	agggtaacat	caattcaaga	aggagggaaa	5160
cacttccttt	ttctggccct	gataatagta	tgagggtgaa	gccaaaataa	aggattcgcg	5220
cccaaatecg	catctttaaa	tgcaggatat	cgatagtctc	tcactctttc	cttactcaag	5280
agtaattctt	gcaaattgct	attatgcaga	tgttataata	tctgtgcgtc	ttgagttgaa	5340
gagctcgaga	ctagtggatc	ccccgggctg	caggaattcg	atatcaagct	tatcgatacc	5400
gtcgacctcg	agggggggcc	cggtagccag	cttttggttc	ctttagttag	ggttaatttc	5460
gagcttggcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	5520
tcacacaaac	ataggagccg	gaagcataaa	gtgtaaagcc	tggggtgcct	aatgagttag	5580
gtaactcaca	ttaattgctg	tgcgctcact	gcccgcttct	cagtcgggaa	acctgtcgtg	5640
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggg	ggtttgcgta	ttgggcgctc	5700
ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcggt	cggtgcggc	gagcggtatc	5760
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	5820
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgctg	tgctggcggt	5880
tttccatagg	ctcgcccccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	5940
gcgaaacccg	acaggactat	aaagatacca	ggcgttcccc	cctggaagct	ccctcgtgcg	6000
ctctcctggt	ccgacctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	6060
cgtggcgctt	tctcaatgct	cacgctgtag	gtatctcagt	tcggtgtagg	tcgttcgctc	6120
caagctgggc	tgtgtgcacg	aacccccctg	tcagcccgac	cgctgcgcct	tatccggtaa	6180
ctatcgtctt	gagtccaacc	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	6240
taacaggatt	agcagagcga	ggtatgtagg	cgggtctaca	gagttcttga	agtgggtggc	6300
taactacggc	tacactagaa	ggacagtatt	tggatatctgc	gctctgctga	agccagttac	6360
cttcggaaaa	agagttggta	gctcttgatc	cggcaaaaca	accaccgctg	gtagcggtag	6420
tttttttggt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	6480
gatcttttct	acggggtctg	acgctcagtg	gaacgaaaac	tcacgttaag	ggatttttgt	6540
catgagatta	tcaaaaagga	tcttcacctt	gatcctttta	aattaaaaat	gaagttttaa	6600
atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	taccaatgct	taatcagtga	6660
ggcacctatc	tcagogatct	gtctatttct	ttcatccata	gttgccctgac	tgcccgctcg	6720
gtagataact	acgatacggg	agggttacc	atctggcccc	agtgtgcaa	tgataccgcg	6780
agaccacgc	tcaccggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	6840
gcgcagaagt	ggtcctgcaa	ctttatccgc	ctccatccag	tctattaatt	gttgccggga	6900
agctagagta	agttagtctc	cagttaatat	tttgcgcaac	gttggttgcca	ttgctacagg	6960
catcgtgggt	tcacgctcgt	cgtttgggtat	ggcttcattc	agctcgggtt	cccaacgata	7020
aaggcgagtt	acatgatccc	ccatgttggt	aaaaaaagcg	gttagctcct	tcggctcctcc	7080
gatcgttggt	agaagtaagt	tggccgcagt	gttatcactc	atgggttatgg	cagcactgca	7140
taattctctt	actgtcatgc	catccgtaag	atgcttttct	gtgactgggt	agtactcaac	7200
caagtcattc	tgagaatagt	gtatgcggcg	accgagttgc	tcttgcccg	cgtcaataacg	7260
ggataatacc	gcgccacata	gcagaacttt	aaaagtgtct	atcattggaa	aacgttcttc	7320

ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	agttcgatgt	aacccactcg	7380
tgcacccaac	tgatcttcag	catcttttac	tttcaccagc	gtttctgggt	gagcaaaaac	7440
aggaaggcaa	aatgccgcaa	aaaagggaat	aagggcgaca	cggaaatgtt	gaatactcat	7500
actcttcctt	tttcaatatt	attgaagcat	ttatcagggt	tattgtctca	tgagcggata	7560
catatttgaa	tgtatttaga	aaaataaaca	aatagggggt	ccgcgcacat	ttccccgaaa	7620
agtgccacct	gacgtcttat	tatcatgaca	ttaacctata	aaaataggcg	tatcacgagg	7680
ccctttcgtc						7690

<210> 42

<211> 9416

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN595

<400> 42

gacgaaaggg	cctcgtgata	cgcctatttt	tatagggttaa	tgatcatgata	ataagacgtc	60
aggtggcact	tttcggggaa	atgtgcgcgg	aacccctatt	tgtttatttt	tctaaataca	120
ttcaaatatg	tatccgctca	tgagacaata	accctgataa	atgcttcaat	aatattgaaa	180
aaggaagagt	atgagtattc	aacatttcgg	tgctgccttt	attccctttt	ttgcggcatt	240
ttgccttcct	gtttttgctc	acccagaaac	gctgggtgaa	gtaaaagatg	ctgaagatca	300
gttggttgca	cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	360
ttttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	420
ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggg	cgcgcgatac	actattctca	480
gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttaaggatg	gcatgacagt	540
aagagaatta	tgacgtgctg	ccataaccat	gagtataaac	actgcggcca	acttacttct	600
gacaacgata	ggaggaccga	aggagctaac	cgtttttttt	cacaacatgg	gggatcatgt	660
aactcgcctt	gacgttggtg	aaccggagct	gaatgaagcc	ataccaaaac	acgagcgtga	720
caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	780
tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	840
acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccggtga	900
gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt	960
agttatctac	acgacgggca	gtcaggcaac	tatggatgaa	cgaatatagac	agatcgtgta	1020
gataggtgcc	tcactgatta	agcattggta	actgtcagac	caagtttact	catatatact	1080
ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	1140
taatctcatg	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccg	1200
agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	gctgcttgca	1260
aacaaaaaaa	ccaccgctac	cagcgggtgg	ttgtttgccg	gatcaagagc	taccaactct	1320
ttttccgaag	gtaactgget	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgtg	1380
gccgtagtta	ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	tcgctctgct	1440
aatcctgtta	ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggactc	1500
aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggt	cgtgcacaca	1560
gccagccttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	agcattgaga	1620
aagcgcacag	cttcccgaag	ggagaaaggc	ggacagggtat	ccggtaagcg	gcagggtcgg	1680
aacaggagag	cgcacgaggg	agcttccagg	ggggaacgcc	tggtatcttt	atagtcctgt	1740
cgggtttcgc	cacctctgac	ttgagcgtcg	atttttgtga	tgctcgtcag	gggggcccag	1800
cctatggaaa	aacgccagca	acgcggcctt	tttacgggtc	ctggcctttt	gctggccttt	1860
tgtcacatg	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	1920
tgagttagct	gataccgctc	gccgcagccg	aacgaccgag	cgcagcaggt	cagtgagcga	1980
ggaagcggaa	gagcgcacca	tacgcaaacc	gcctctcccc	gcgcgttgge	cgattcatta	2040
atgcagctgg	cacgacaggt	ttcccgaactg	gaaagcgggc	agttagcgcga	acgcaattaa	2100
tgtgagttac	ctcactcatt	aggcacccca	ggctttacac	tttatgcttc	cggctcctat	2160
gttgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	2220
cgccaagctc	ggaattaacc	ctcactaaag	ggaacaaaag	ctgggtaccg	ggccccccct	2280

cgagggtcgac	ggtatcgata	agcttgatat	cgaattcctg	cagccccggg	gatccactag	2340
tctcgagctc	ttcaactcaa	gacgcacaga	tattataaca	tctgcataat	aggcatttgc	2400
aagaattact	cgtgagtaag	gaaagagtga	ggaactatcg	catacctgca	tttaaagatg	2460
ccgatttggg	cgcgaatcct	ttattttggc	ttcaccctca	tactattatc	agggccagaa	2520
aaaggaagtg	tttccctcct	tcttgaattg	atgttaccct	cataaagcac	gtggcctctt	2580
atcgagaaag	aaattaccgt	cgctcgtgat	ttgtttgcaa	aaagaacaaa	actgaaaaaa	2640
cccagacacg	ctcgacttcc	tgtcttccta	ttgattgcag	cttccaattt	cgtcacacaa	2700
caaggctcta	gcgacggctc	acaggttttg	taacaagcaa	tcgaaggttc	tggaatggcg	2760
ggaaaggggt	tagtaccaca	tgctatgatg	cccactgtga	tctccagagc	aaagttcggt	2820
cgatcgtact	gttactctct	ctctttcaaa	cagaattgtc	cgaatcgtgt	gacaacaaca	2880
gctgtttctc	acacactctt	ttcttctaac	caaggggggtg	gttttagttta	gtagaacctc	2940
gtgaaactta	cattttacata	tatataaaact	tgcataaatt	ggtcaatgca	agaaatacat	3000
atttgggtctt	ttctaattcg	tagtttttca	agttcttaga	tgctttcttt	ttctcttttt	3060
tacagatcat	caaggaagta	attatctact	ttttacaaca	aatatatcta	gaaaatggct	3120
gttaccaacg	ttgctgaatt	gaacgctttg	gttgaaaggg	ttaagaaggc	tcaaagagaa	3180
tacgcttctt	tcaccaaga	acaagttgac	aagatcttca	gagctgctgc	tttggtgct	3240
gctgacgcta	gaatcccat	ggctaagatg	gctgttgctg	aatctgggtat	gggtatcggt	3300
gaagacaagg	ttatcaagaa	ccacttcgct	tctgaataca	tctacaacgc	ttacaaggac	3360
gaaaagacct	gtgggtgttt	gtcagaagac	gacaccttcg	gtaccatcac	catcgtgaa	3420
ccaatcggta	tcctctgtgg	tatcgttcca	accaccaacc	caacctctac	cgtatcttct	3480
aagtctttga	tctctttgaa	gaccagaaac	gctatcatct	tctctccaca	cccaagagct	3540
aaagacgcta	ccaacaaggc	tgctgacatc	gttttgcaag	ctgctatcgc	tgctgggtgct	3600
ccaaaggact	tgatcggttg	gatcgaccaa	ccatctgttg	aattgtctaa	cgttttgatg	3660
caccacccag	acatcaactt	gatcttggct	accggtggct	caggtatggg	taaggctgct	3720
tactcttctg	gtaagccagc	tatcgggtgt	gggtgctggt	acaccccggt	tggtatcgac	3780
gaaaccgctg	acatcaagag	agctgttgct	tctgttttga	tgtctaagac	cttcgacaac	3840
ggtgttatct	gtgcttctga	acaatctgtt	gttgttggtg	actctgttta	cgacgctgtt	3900
agagaaagat	tcgctaccca	cgggtggttac	ttgttgcaag	gtaaggaatt	gaaggctgtt	3960
caagacgtta	tcttgaagaa	cgggtgctttg	aacgctgcta	tcgttggtca	accagcttac	4020
aagatcgtctg	aattagctgg	tttctctgtt	ccagaaaaca	ccaagatctt	gatcgggtgaa	4080
gttaccgttg	ttgacgaatc	tgaaccatct	gctcacgaaa	agttgtctcc	aaccttggt	4140
atgtacagag	ctaaggactt	cgaagacgct	gttgaaaaag	ctgaaaagtt	ggttgctatg	4200
ggtgggtattg	gtcacacctc	ttgtttgtac	accgaccaag	acaaccaacc	agctagagtt	4260
tcttacttctg	gtcaaaaagat	gaagaccgct	agaatcttga	tcaacacccc	agcttctcaa	4320
ggtgggtatcg	gtgacttgta	caacttcaag	ttggtcccat	ctttgacctt	gggttggtgt	4380
tcttgggggtg	gtaactctat	ctctgaaaac	gttgggtccaa	agcacttgat	caacaagaag	4440
accgttgcta	agagagctga	aaacatgttg	tggcacaagt	tgccaaaatc	tatctacttc	4500
agaagagggtt	ctttgccaat	cgttttggtg	gaagttatca	ccgacgggtca	caagagagct	4560
ttgatcgtta	ccgacagatt	cttggttcaac	aacggttacg	ctgaccaaata	cacctctgtt	4620
ttgaaggctg	ctgggtgttg	aaccgaagtt	ttcttcgaag	ttgaagctga	cccaaccttg	4680
tctatcgtta	gaaaggggtg	tgaattgggt	aactctttca	agccagacgt	tatcatcgct	4740
ttgggtgggtg	gttctccaat	ggacgctgct	aagatcatgt	gggttatgta	cgaacaccca	4800
gaaacccact	tcgaagaatt	ggctttgaga	ttcatggaca	tcagaaagag	aatctacaag	4860
ttcccaaaga	tgggtgttaa	ggctaagatg	atcgtgtgta	ccaccacctc	tgggtaccggt	4920
tctgaagtta	ccccattcgc	tggttggtacc	gacgacgcta	ccggtcaaaa	gtaccatttg	4980
gctgactacg	ctttgacccc	agacatgggt	atcgttgacg	ctaacttggg	tatggacatg	5040
ccaaagtctt	tgtgtgcttt	cgggtggtttg	gacgctgtta	cccacgctat	ggaagcttac	5100
gtttctgttt	tggttctctga	attctctgac	ggtcaagctt	tgcaagcttt	gaagttgttg	5160
aaggaatact	tgccagcttc	ttaccacgaa	ggttctaaga	accagtttgc	tagagaaaga	5220
gttcactctg	ctgctaccat	cgttggtatc	gctttcgtca	acgctttctt	gggtgtttgt	5280
cactctatgg	ctcacaagtt	gggttctcaa	ttccacatcc	cacacggttt	ggctaacgct	5340
ttgttgatct	gtaacgttat	cagatacaac	gctaacgaca	acccaacca	gcaaaccgct	5400
ttctctcaat	acgacagacc	acaagctaga	agaagatacg	ctgaaatcgc	tgaccacttg	5460
ggtttgctctg	ctccaggtga	cagaacogct	gcaaagatcg	aaaagttgtt	ggcttggttg	5520
gaaaccttga	aggctgaatt	gggtatccca	aagtctatca	gagaagctgg	tgttcaagaa	5580
gctgacttct	tggctaacgt	tgacaagttg	tctgaagacg	ctttcgacga	ccaatgtacc	5640

ggtgctaacc	caagataccc	attgatctct	gaattgaago	aaatcttggt	ggacacctac	5700
tacggtagag	actacgttga	aggtgaaacc	gctgctaaga	aggaagctgc	tccagctaag	5760
gctgaaaaga	aggctaagaa	gtctgcttag	cttaagcgcg	cgaatttctt	atgatttatg	5820
atTTTTatta	ttaaataagt	tataaaaaaa	ataagtgtat	acaaatttta	aagtgactct	5880
taggttttaa	aacgaaaatt	cttattcttg	agtaactctt	tcctgtaggt	caggttgctt	5940
tctcaggtat	agcatgaggt	cgctcttatt	gaccacacct	ctaccggcat	gccgagcaaa	6000
tgcctgcaaa	tcgctcccca	tttcacccaa	ttgtagatat	gctaaactcca	gcaatgagtt	6060
gatgaatctc	ggtgtgtatt	ttatgtcctc	agaggacaac	acctgttgta	atcgttcttc	6120
cacacgtacg	aagcttaaaa	gggcgaattc	tgcatatato	catcacactg	gcggcgcat	6180
gctagctccg	gattatcgat	gataagctgt	caaagatgag	aattaattcc	acggactata	6240
gactatacta	gatactccgt	ctactgtacg	atacacttcc	gctcaggtec	ttgtccttta	6300
acgaggcctt	accactcttt	tgttactcta	ttgatccagc	tcagcaaagg	cagtgtgatc	6360
taagattcta	tcttcgcat	gtagtaaaac	tagctagacc	gagaaagaga	ctagaaatgc	6420
aaaaggcact	tctacaatgg	ctgccatcat	tattatccga	tgtgacgctg	cagcttctca	6480
atgatattcg	aatacgcttt	gaggagatac	agcctaatat	ccgacaaact	gttttacaga	6540
tttacgatcg	tacttgttac	ccatcattga	atTTTgaaca	tccgaacctg	ggagttttcc	6600
ctgaaacaga	tagtataatt	gaacctgtat	aataatatat	agtctagcgc	tttacggaag	6660
acaatgtatg	tatttcggtt	cctggagaaa	ctattgcate	tattgcatag	gtaatcttgc	6720
acgtcgcate	cccggttcat	tttctgcgtt	tccatcttgc	acttcaatag	catatctttg	6780
ttaacgaage	atctgtgctt	cattttgtag	aacaaaaatg	caacgcgaga	gcgctaattt	6840
ttcaaacaaa	gaatctgagc	tgcatTTTTa	cagaacagaa	atgcaacgcg	aaagcgctat	6900
tttaccaacg	aagaatctgt	gcttcatttt	tgtaaaacaa	aaatgcaacg	cgacgagagc	6960
gctaattttt	caaacaaaga	atctgagctg	cattttttaca	gaacagaaat	gcaacgcgag	7020
agcgtatatt	taccaacaaa	gaatctatac	ttctTTTTtg	ttctacaaaa	atgcatcccg	7080
agagcgctat	ttttctaaca	aagcatctta	gattactttt	tttctccttt	gtgcgctcta	7140
taatgcagtc	tcttgataac	tttttgcaact	gtaggtccgt	taaggttaga	agaaggctac	7200
tttggtgtct	atTTTctctt	ccataaaaaa	agcctgactc	cacttcccgc	gtttactgat	7260
tactagcgaa	gctgcgggtg	catttttttca	agataaaaggc	atccccgatt	atattctata	7320
ccgatgtgga	ttgcgcatac	tttgTgaaca	gaaagtgata	gcgttgatga	ttcttcattg	7380
gtcagaaaaat	tatgaacggg	ttcttctatt	ttgtctctat	atactacgta	taggaaatgt	7440
ttacattttc	gtattgtttt	cgattcactc	tatgaatagt	tcttactaca	atTTTTttgt	7500
ctaaagagta	atactagaga	taaacataaa	aaatgtagag	gtcgagttaa	gatgcaagtt	7560
caaggagcga	aagggtggatg	ggtagggttat	atagggatat	agcacagaga	tatatagcaa	7620
agagatactt	ttgagcaatg	tttgTggaag	cggtattcgc	aatgccggct	ttccccgtca	7680
agctctaaat	cgggggctcc	cttttaggggt	ccgatttagt	gctttacggc	acctcgaccc	7740
caaaaaactt	gattaggggtg	atggttcacg	tagtgggcca	tcgccctgat	agacggtttt	7800
tcgccctttg	acgttgaggt	ccacgttctt	taatagtgga	ctcttgttcc	aaactggaac	7860
aacactcaac	cctatctcgg	tctattcttt	tgatttataa	gggattttgc	cgatttcggc	7920
ctattgggtta	aaaaatgagc	tgatttaaca	aaaattttaac	gcgaatttta	acaaaatatt	7980
aacgtttaca	atTTTctgat	gcggtatTTT	ctccttacgc	atctgtgcgg	tatttcacac	8040
cgcataatgat	ccgtcaggtt	caagagaaaa	aaaaagaaaa	agcaaaaaaga	aaaaaggaaa	8100
gcgcgcctcg	ttcagaatga	caagtataga	atgatgcatt	accttggtcat	cttcagtatc	8160
atactgttcg	tatacatact	tactgacatt	cataggtata	catatataca	catgtatata	8220
tatcgtatgc	tgcatcttta	aataatcggt	gtcactacat	aagaacacct	ttggtggagg	8280
gaacatcggt	ggtaccattg	ggcgagggtg	cttctcttat	ggcaaccgca	agagccttga	8340
acgcactctc	actacgggtga	tgatcattct	tgccctgcag	acaatcaacg	tggagggtaa	8400
ttctgctage	ctctgcaaag	ctttcaagaa	aatgcgggat	catctcgcaa	gagagatctc	8460
ctactttctc	cctttgcaaa	ccaagtctga	caactgcgta	cggcctgttc	gaaagatcta	8520
ccaccgctct	ggaaagtgcc	tcatccaaag	gcgcaaatcc	tgatccaaac	ctttttactc	8580
cacgcgccag	tagggcctct	ttaaaagctt	gaccgagagc	aatccccgag	tcttcagtgg	8640
tgtgatggtc	gtctatgtgt	aagtcaccaa	tgactcaac	gattagcgac	cagccggaat	8700
gcttggccag	agcatgtatc	atatgggtcca	gaaaccctat	acctgtgtgg	acgttaatca	8760
cttgcgattg	tgtggcctgt	tctgctactg	cttctgcctc	ttttctctggg	aagatcgagt	8820
gctctatcgc	taggggacca	ccctttaaag	agatcgcaat	ctgaatcttg	gtttcatttg	8880
taatacgctt	tactagggct	ttctgctctg	tcactctttg	cttcgtttat	cttgccctgct	8940
catttttttag	tatattcttc	gaagaaatca	cattacttta	tataatgtat	aattcattat	9000

gtgataatgc	caatcgctaa	gaaaaaaaaa	gagtcacccg	ctaggtggaa	aaaaaaaaat	9060
gaaaatcatt	accgagcat	aaaaaaaaat	agagtgtact	agaggaggcc	aagagtaata	9120
gaaaaagaaa	attgcgggaa	aggactgtgt	tatgacttcc	ctgactaatg	ccgtgttcaa	9180
acgataacctg	gcagtgactc	ctagcgctca	ccaagctctt	aaaacggaat	tatggtgcac	9240
tctcagtaca	atctgctctg	atgccgcata	gttaagccag	ccccgacacc	cgccaacacc	9300
cgctgacgcg	ccctgacggg	cttgtctgct	cccggcatcc	gcttacagac	aagctgtgac	9360
cgtctccggg	agctgcatgt	gtcagagggt	ttcaccgtca	tcaccgaaac	gcgcga	9416

<210> 43

<211> 9352

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN596

<400> 43

gacgaaaggg	cctcgtgata	cgcctatttt	tatagggttaa	tgctcatgata	ataagacgtc	60
aggtggcact	tttcggggaa	atgtgcgcgg	aacccttatt	tgtttatttt	tctaaataca	120
ttcaaataatg	tatccgctca	tgagacaata	accctgataa	atgcttcaat	aatattgaaa	180
aaggaagagt	atgagtattc	aacatttccg	tgctgccttt	attccctttt	ttgcggcatt	240
ttgccttcc	gtttttgctc	accagaaac	gctgggtgaa	gtaaaagatg	ctgaagatca	300
gttgggtgca	cgagtgggtt	acatcgaa	ggatctcaac	agcggtaaga	tccttgagag	360
tttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	420
ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcgg	cgccgcatac	actattctca	480
gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	540
aagagaatta	tgacgtgctg	ccataaccat	gagtggataa	actgcggcca	acttacttct	600
gacaacgatc	ggaggaccga	aggagctaac	cgcttttttt	cacaacatgg	gggatcatgt	660
aactcgctt	gatcggtggg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	720
caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	780
tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	840
acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccggtga	900
gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgatcgt	960
agttatctac	acgacgggca	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	1020
gatagggtgcc	tactgatta	agcattggta	actgtcagac	caagtttact	catatatact	1080
ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	1140
taatctcatg	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccgt	1200
agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	gctgcttgca	1260
aacaaaaaaa	ccaccgctac	cagcgggtgg	ttgtttgccg	gatcaagagc	taccaactct	1320
ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgtg	1380
gccgtagtta	ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	tcgctctgct	1440
aatcctgtta	ccagtggctg	ctgccagtg	cgataagtcg	tgtcttaccg	ggttggaactc	1500
aagacgatag	ttaccgggata	aggcgcagcg	gtcgggctga	acgggggggt	cgtgcacaca	1560
gccagcttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	agcattgaga	1620
aagcggcacg	cttcccgaag	ggagaaaggc	ggacaggtat	ccggtaagcg	gcagggtcgg	1680
aacaggagag	cgcacgaggg	agcttccagg	ggggaacgcc	tggtatcttt	atagtcctgt	1740
cgggtttcgc	cacctctgac	ttgagcgctg	atttttgtga	tgctcgtcag	gggggcccag	1800
cctatggaaa	aacgccagca	acgcggcctt	tttacgggtc	ctggcctttt	gctggccttt	1860
tgctcacatg	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	1920
tgagtgagct	gataccgctc	gccgcagccg	aacgaccgag	cgcagcaggt	cagtgagcga	1980
ggaagcggaa	gagcgcccaa	tacgcaaacc	gcctctcccc	gcgcgttggc	cgattcatta	2040
atgcagctgg	cacgacaggt	ttcccgaactg	gaaagcgggc	agtgagcgca	acgcaattaa	2100
tgtgagttac	ctcactcatt	aggcacccca	ggcttttacac	tttatgcttc	cggctcctat	2160
gttgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	2220
cgccaagctc	ggaattaacc	ctcactaaag	ggaacaaaag	ctgggtaccg	ggccccccct	2280

cgagggtcgac	ggtatcgata	agcttgatat	cgaattcctg	cagcccgggg	gatccactag	2340
tctcgagctc	ttcaactcaa	gacgcacaga	tattataaca	tctgcataat	aggcatttgc	2400
aagaattact	cgtgagtaag	gaaagagtga	ggaactatcg	catacctgca	tttaaagatg	2460
ccgatttggg	cgcgaatcct	ttattttggc	ttcaccctca	tactattatc	agggccagaa	2520
aaaggaagtg	tttccctcct	tcttgaattg	atgttaccct	cataaagcac	gtggcctctt	2580
atcgagaaag	aaattaccgt	cgctcgtgat	ttgtttgcaa	aaagaacaaa	actgaaaaaa	2640
cccagacacg	ctcgacttcc	tgtcttccct	ttgattgcag	cttccaattt	cgtcacacaa	2700
caagggtccta	gcgacggctc	acagggtttt	taacaagcaa	tcgaagggtc	tggaatggcg	2760
ggaaaggggt	tagtaccaca	tgctatgatg	cccactgtga	tctccagagc	aaagttcgtt	2820
cgatcgact	gttactctct	ctcttttcaa	cagaattgtc	cgaatcgtgt	gacaacaaca	2880
gcctgttctc	acacactctt	ttctttctaac	caaggggggt	gtttagttta	gtagaacctc	2940
gtgaaactta	catttacata	tatataaact	tgcataaatt	ggtcaatgca	agaaatacat	3000
atttgggtctt	ttctaattcg	tagtttttca	agttcttaga	tgctttcttt	ttctcttttt	3060
tacagatcat	caaggaagta	attatctact	ttttacaaca	aatatatcta	gaaaatgtct	3120
acccaacaaa	ccatgaccgt	tgacgaacac	atcaaccaat	tagttagaaa	ggctcaagtt	3180
gctttgaagg	aatacttgaa	gccagaatac	acccaagaaa	agatcgacta	catcgttaag	3240
aaggcttctg	ttgctgcttt	ggaccaacac	tgtgcttttg	ctgctgctgc	tgttgaagaa	3300
accggtagag	gtatcttcga	agacaaggct	accaagaaca	tcttcgcttg	tgaacacggt	3360
accacgaaa	tgagacacgc	taagaccggt	ggtatcatca	acgttgaccc	attgtacggt	3420
atcaccgaaa	tcgctgaacc	agttgggtgt	gtttgtgggt	ttaccccagt	taccaacca	3480
acctctaccg	ctatcttcaa	gtctttgate	tctatcaaga	ccagaaaccc	aatcgttttc	3540
tctttccacc	catctgcttt	gaagtgttct	attatggctg	ctaaaatcgt	tagagacgct	3600
gctatcgctg	ctgggtgctcc	agaaaactgt	atccaatgga	tcgaattcgg	tggtatcgaa	3660
gcttctaaca	agttgatgaa	ccaccacagg	gttgctacta	tcttggctac	cgggtggtaac	3720
gctatggtta	aggctgcata	ctcttctggg	aagccagctt	tgggtgttgg	tgctggtaac	3780
gttccaacct	acatcgaaaa	gacctgtaac	atcaagcaag	ctgctaacga	cgttgttatg	3840
tctaagtctt	tcgacaacgg	tatgatctgt	gcttctgaac	aagctgctat	catcgacaag	3900
gaaatctacg	accaagttgt	tgaagaaatg	aagaccttgg	gtgcttactt	catcaacgaa	3960
gaagaaaagg	ctaagttgga	aaagttcatg	ttcgggtgta	acgcttactc	tgctgacggt	4020
aacaacgcta	gattgaaccc	aaagtgtcca	ggtatgtctc	cacaatgggt	cgtgaaacaa	4080
gttgggtatca	aggtaccaga	agactgtaac	atcatctgtg	ctgtttgtaa	ggaagttggg	4140
ccaaacgaac	cattgaccag	agaaaagttg	tctccagttt	tggctatctt	gaaagctgaa	4200
aacacccaag	acggtatcga	caaggctgaa	gctatgggtg	aatttaacgg	tagaggtcac	4260
tctgctgcta	tccactctaa	cgacaaggct	gttggtgaaa	agtacgcttt	gaccatgaag	4320
gcttgtagaa	tcttgacaaa	caccccatct	tctcaagggt	gtatcggttc	tatctacaac	4380
tacatctggc	catctttcac	cttgggttgt	ggttcttacg	gtggtaactc	tgtttctgct	4440
aacgttacct	accacaactt	gttgaacatc	aagagattgg	ctgacagaag	aaacaacttg	4500
caatgggtca	gagttccacc	aaagatcttc	ttcgaaccac	actctatcag	atacttgggt	4560
gaattgaagg	aattgtctaa	gatcttcac	gtttctgaca	gaatgatgta	caagttgggt	4620
tacgttgaca	gagttatgga	cgttttgaag	agaagatcta	acgaagttga	aatcgaaatc	4680
ttcatcgacg	ttgaaccaga	cccatctatc	caaaccgttc	aaaagggttt	ggctgttatg	4740
aacaccttcg	gtccagacaa	catcatcgct	atcgggtggg	gttctgctat	ggacgctgct	4800
aagatcatgt	ggttggttga	cgaacaccca	gaagctgact	tcttcgctat	gaagcaaaaag	4860
ttcatcgact	tgagaaagag	agctttcaag	ttcccaacca	tgggtaagaa	ggctagattg	4920
atctgtatcc	caaccacctc	tgggtaccgt	tctgaagtta	ccccattcgc	tgttatctct	4980
gaccacgaaa	ccggtaaagaa	gtacccattg	gctgactact	ctttgacccc	atctgttgct	5040
atcgttgacc	caatgttcac	catgtctttg	ccaaagagag	ctatcgctga	caccggtttg	5100
gacgttttgg	ttcacgctac	cgaagcttac	gtttctgtta	tggctaacga	atacaccgac	5160
ggtttggcta	gagaagctgt	taagttgggt	tttgaaaact	tgttgaagtc	ttacaacggt	5220
gacttggaag	ctagagaaaa	gatgcacaac	gctgctacca	tcgctgggat	ggctttcgct	5280
tctgctttct	tgggtatgga	ccactctatg	gctcacaagg	ttgggtgctgc	tttccacttg	5340
ccacacggta	gatgtgttgc	tgttttgttg	ccacacgtta	tcagatacaa	cgggtcaaaaag	5400
ccaagaaagt	tggctatgtg	gccaaagtac	aacttctaca	aggctgacca	aagatacatg	5460
gaattggctc	aaatggttgg	tttgaagtgt	aacaccccag	ctgaaggtgt	tgaagctttc	5520
gctaaggctt	gtgaagaatt	gatgaaggct	accgaaacca	tcaccggttt	caagaaggct	5580
aacatcgacg	aagctgcttg	gatgtctaa	gttccagaaa	tggctttggt	ggctttcgaa	5640

gaccaatggt	ctccagctaa	cccaagaggt	ccaatgggtta	aggacatgga	aaagatcttg	5700
aaggetgctt	actacccaat	cgcttagctt	aagcgcgcga	atttcttatg	atttatgatt	5760
tttattatta	aataagttat	aaaaaaaaata	agtgtataca	aatttttaaag	tgactcttag	5820
gtttttaaac	gaaaattctt	attcttgagt	aactctttcc	tgtaggtcag	gttgctttct	5880
caggtatagc	atgaggtcgc	tcttattgac	cacacctcta	ccggcatgcc	gagcaaatgc	5940
ctgcaaatcg	ctccccattt	cacccaattg	tagatatgct	aactccagca	atgagttgat	6000
gaatctcggt	gtgtatttta	tgtcctcaga	ggacaacacc	tgttgtaatc	gttcttccac	6060
acgtacgaag	ctaaaagggc	gaattctgca	gatataccatc	acactggcgg	ccgcatgcta	6120
gctccggatt	atcgatgata	agctgtcaaa	gatgagaatt	aattccacgg	actatagact	6180
atactagata	ctccgtctac	tgtacgatac	acttccgctc	aggtccttgt	cctttaacga	6240
ggccttacca	ctcttttggt	actctattga	tccagctcag	caaaggcagt	gtgatctaag	6300
attctatctt	cgcgatgtag	taaaactagc	tagaccgaga	aagagactag	aaatgcaaaa	6360
ggcacttcta	caatggctgc	catcattatt	atccgatgtg	acgctgcagc	ttctcaatga	6420
tattcgaata	cgctttgagg	agatacagcc	taatatccga	caaactgttt	tacagattta	6480
cgatcg tact	tgttaccat	cattgaattt	tgaacatccg	aacctgggag	ttttccctga	6540
aacagatagt	atatttgaac	ctgtataata	atatatagtc	tagcgcctta	cggaagacaa	6600
tgtatgtatt	tccggttctg	gagaaactat	tgcactctatt	gcataggtaa	tcttgcacgt	6660
cgcacccccg	gttcattttc	tgcgtttcca	tcttgcaactt	caatagcata	tctttgttaa	6720
cgaagcatct	gtgcttcatt	ttgtagaaca	aaaatgcaac	gcgagagcgc	taatttttca	6780
aacaaagaat	ctgagctgca	tttttacaga	acagaaatgc	aacgcgaaag	cgctatttta	6840
ccaacgaaga	atctgtgctt	catttttgta	aaacaaaaat	gcaacgcgac	gagagcgcta	6900
atttttcaaa	caaagaatct	gagctgcatt	tttacagaac	agaaatgcaa	cgcgagagcg	6960
ctattttacc	aacaaagaat	ctatacttct	tttttggtct	acaaaaatgc	atcccgagag	7020
cgctattttt	ctaacaaagc	atcttagatt	actttttttc	tcctttgtgc	gctctataat	7080
gcagtctctt	gataactttt	tgcactgtag	gtccgttaag	gttagaagaa	ggctactttg	7140
gtgtctattt	tctcttccat	aaaaaaagcc	tgactccaact	tcccgcgttt	actgattact	7200
agcgaagctg	cgggtgcatt	ttttcaagat	aaaggcatcc	ccgattatat	tctataccga	7260
tgtggattgc	gcatactttg	tgaacagaaa	gtgatagcgt	tgatgattct	tcattgggtca	7320
gaaaattatg	aacggtttct	tctattttgt	ctctatatac	tacgtatagg	aaatgtttac	7380
attttctgat	tgttttcgat	tcactctatg	aatagttctt	actacaattt	ttttgtctaa	7440
agagtaatac	tagagataaa	cataaaaaat	gtagaggctg	agtttagatg	caagttcaag	7500
gagcgaaagg	tggatgggta	ggttatatag	ggatatagca	cagagatata	tagcaaagag	7560
atacttttga	gcaatgtttg	tggaaagcgt	attcgcaatg	ccggctttcc	ccgtcaagct	7620
ctaaatcggg	ggctcccttt	agggttccga	tttagtgctt	tacggcacct	cgaccccaaa	7680
aaacttgatt	agggatgatg	ttcacgtagt	gggccatcgc	cctgatagac	ggtttttctgc	7740
cctttgacgt	tggagtccac	gttctttaat	agtggactct	tgttccaaac	tggaaacaaca	7800
ctcaacccta	tctcgggtcta	ttcttttgat	ttataaggga	ttttgccgat	ttcggcctat	7860
tgggttaaaaa	atgagctgat	tttaacaaaa	tttaacgcga	atttttaacaa	aatattaacg	7920
tttacaattt	cctgatgcgg	tattttctcc	ttacgcactct	gtgcgggtatt	tcacaccgca	7980
tatgatccgt	cgagttcaag	agaaaaaaaa	agaaaaagca	aaaagaaaaa	aggaaagcgc	8040
gcctcggttca	gaatgacacg	tatagaatga	tgcattacct	tgtcatcttc	agtatcatac	8100
tgttcgtata	catacttact	gacattcata	ggatatacata	tatacacatg	tatataatc	8160
gtatgctgca	gcttttaaata	atcgggtgtca	ctacataaga	acaccttttg	tggagggaac	8220
atcgttggtg	ccattgggcg	aggtggcttc	tcttatggca	accgcaagag	ccttgaacgc	8280
actctcacta	cggatgatgat	cattcttgcc	tccgagacaa	tcaacgtgga	gggtaattct	8340
gctagcctct	gcaaagcttt	caagaaaatg	cgggatcctc	tccgaagaga	gatctcctac	8400
tttctccctt	tgcaaaccac	gttcgacaac	tgcgtacggc	ctgttcgaaa	gatctaccac	8460
cgcctctggaa	agtgcctcat	ccaaaggcgc	aaatcctgat	ccaaaccttt	ttactccacg	8520
cgcagtagg	gcctctttta	aagcttgacc	gagagcaatc	cgcagctctt	cagtgggtgtg	8580
atggctcgtc	atgtgtaagt	caccaatgca	ctcaacgatt	agcgaccagc	cggaatgctt	8640
ggccagagca	tgtatcatat	ggtccagaaa	ccctatacct	gtgtggacgt	taatcacttg	8700
cgattgtgtg	gcctgtttctg	ctactgcttc	tgcctctttt	tctgggaaga	tcgagtgtctc	8760
tatcgttagg	ggaccaccct	ttaaagagat	cgcgaatctga	atcttggttt	catttgtaat	8820
acgctttact	agggctttct	gctctgtcat	ctttgccttc	gtttatcttg	cctgctcatt	8880
ttttagttata	ttcttcgaag	aaatcacatt	acttttatata	atgtataatt	cattatgtga	8940
taatgccaat	cgctaagaaa	aaaaaagagt	catccgctag	gtggaaaaaa	aaaaatgaaa	9000

atcattaccg aggcataaaa aaatatagag tgtactagag gaggccaaaga gtaatagaaa 9060
aagaaaattg cgggaaagga ctgtgttatg acttccctga ctaatgccgt gttcaaacga 9120
tacctggcag tgactcctag cgctcaccaa gctcttaaaa cggaattatg gtgcactctc 9180
agtacaatct gctctgatgc cgcatagtta agccagcccc gacaccgcc aacaccgcc 9240
gacgcgccct gacgggcttg tctgctcccg gcacccgctt acagacaagc tgtgaccgtc 9300
tccgggagct gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc ga 9352

<210> 44
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer GPD2uf

<400> 44
ggtaccagat cttttgcggc gaggtgccg 29

<210> 45
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer GPD2ur

<400> 45
tctagactta aggaatgtgt atcttgtaa tcttctgaca gc 42

<210> 46
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer GPD2df

<400> 46
ctcgagatag tctacaacaa cgtccgca 28

<210> 47
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer GPD2dr

<400> 47
ccatggagat ctgcagtga aaagctcgaa gaaacagct 39

<210> 48
 <211> 4397
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pRN594

<400> 48

gacgacggcg	acctggcggg	cttcgtggtc	gtctcgtaact	ccggctggaa	ccgccggctg	60
accgtcgagg	acatcgaggt	cgccccggag	caccgggggc	acgggggtcgg	gcgcgcgttg	120
atggggctcg	cgacggagtt	cgccccgcgag	cgggggcgccg	ggcacctctg	gctggagggtc	180
accaacgtca	acgcaccggc	gatccacgcg	taccggcgga	tgggggttcac	cctctgcggc	240
ctggacaccg	ccctgtacga	cggcaccgcc	tcggacggcg	agcaggcgct	ctacatgagc	300
atgccctgcc	cctagtactg	acaataaaaa	gattcttgtt	ttcaagaact	tgtcatttgt	360
atagtttttt	tatattgtag	ttgttctatt	ttaatcaa	gttagcgtga	tttatatttt	420
ttttcgctc	gacatcatct	gcccgatgc	gaagttaagt	gcgcagaaag	taatatcatg	480
cgtcaatcgt	atgtgaatgc	tggctcgctat	actgctgtcg	attcgatact	aacgcgcgca	540
tccagtgtcg	acggatccta	ggtgtacagg	gcccacaaagg	gcgaattctg	cagatatcca	600
tcacactggc	ggccgctcga	gatagtctac	aacaacgtcc	gcatggaaga	cctaccggag	660
atgattgaag	agctagacat	cgatgacgaa	tagacactct	ccccccccct	ccccctctga	720
tctttcctgt	tgctcttttt	tcccccaacc	aatttatcat	tatacacaag	ttctacaact	780
actactagta	acattactac	agttattata	attttctatt	ctctttttct	ttaagaatct	840
atcattaacg	ttaatttcta	tatatacata	actaccatta	tacacgctat	tatcgtttac	900
atatcacatc	accgttaatg	aaagatacga	caccctgtac	actaacacaa	ttaaataatc	960
gccataacct	tttctgttat	ctatagccct	taaagctgtt	tcttcgagct	ttttcactgc	1020
agatctccat	ggcgatgcct	gcttgccgaa	tatcatggtg	gaaaatggcc	gcttttctgg	1080
attcatcgac	tgtggccggc	tgggtgtggc	ggaccgctat	caggacatag	cgttggctac	1140
ccgtgatatt	gctgaagagc	ttggcggcga	atgggctgac	cgcttctctg	tgctttacgg	1200
tatcgccgct	cccgattcgc	agcgcatcgc	cttctatcgc	cttcttgacg	agttcttctg	1260
aattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	tccctttttt	1320
gcggcatttt	gccttctgt	ttttgctcac	ccagaaacgc	tggtgaaagt	aaaagatgct	1380
gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	1440
cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	1500
tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	ccgcatacac	1560
tattctcaga	atgacttgg	tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	1620
atgacagtaa	gagaattatg	cagtgtgtcc	ataaccatga	gtgataaac	tgcgccaac	1680
ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	caacatgggg	1740
gatcatgtaa	ctcgcttga	tcgttgggaa	ccggagctga	atgaagccat	accaaacgac	1800
gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaact	attaactggc	1860
gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	ggataaagtt	1920
gcaggaccac	ttctgcgctc	ggcccttcgc	gctggctgg	ttattgctga	taaatctgga	1980
gccggtgagc	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	taagcctcc	2040
cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	aaatagacag	2100
atcgtgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	agtttactca	2160
tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	2220
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	2280
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	2340
tgcttgcaaa	caaaaaaacc	accgtacca	gcggtggttt	gtttgcccga	tcaagagcta	2400
ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgttctt	2460
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	2520
gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtcgtg	tcttaccggg	2580
ttggactcaa	gacgatagtt	accggataag	gcgcagcgg	cgggctgaac	gggggggttcg	2640
tgcacacagc	ccagcttgg	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	2700
ctatgagaaa	gcgccacgct	tcccgaagg	agaaaggcgg	acaggtatcc	ggtaagcggc	2760
agggctcgaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	2820

```

agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg 2880
gggcgaggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc 2940
tggccttttg ctcacatggt ctttctgctg ttatcccttg attctgtgga taaccgtatt 3000
accgcctttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca 3060
gtgagcgagg aagcggaaga gcgcccaata cgcaaaccgc ctctcccccgc gcgttggccg 3120
attcattaat gcagctggca cgacagggtt cccgactgga aagcgggcag tgagcgcaac 3180
gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg 3240
gctcgtatgt tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac 3300
catgattacg ccaagctatt taggtgacac tatagaatac tcaagctatg catcaagctt 3360
ggtaccagat cttttgcggc gaggtgcga tgggttgctg aggggaagag tgtttagctt 3420
acggacctat tgccattgtt attccgatta atctattgtt cagcagctct tctctacct 3480
gtcattctag tttttttttt tttttttttt ggttttactt ttttttcttc ttgccttttt 3540
ttcttggttac tttttttcta gttttttttc cttccactaa gctttttcct tgatttatcc 3600
ttgggttctt ctttctactc ctttagattt tttttttata tattaatttt taagtttatg 3660
tatttttgta gattcaattc tctttccctt tccttttctt tcgctccctt tccttatcaa 3720
tgcttgctgt cagaagatta acaagataca cattccttaa ggctcgtcc ccgcccgggc 3780
acccggccag cgacatggag gcccagaata ccctccttga cagtcttgac gtgcgcagct 3840
caggggcatg atgtgactgt cgcccgtaac tttagcccat acatcccat gtataatcat 3900
ttgcatccat acattttgat ggccgcacgg cgcgaagcaa aaattacggc tcctcgtgc 3960
agacctgcga gcagggaac gctccctca cagacgcgtt gaattgtccc cacgccgcgc 4020
ccctgtagag aaatataaaa ggtaggatt tgccactgag gttcttcttt catatacttc 4080
cttttaaaat cttgctagga tacagttctc acatcacatc cgaacataaa caaccatgta 4140
aaatgaccac tcttgacgac acggcttacc ggtaccgcac cagtgtcccg ggggacgccg 4200
aggccatcga ggcactggat gggctcttca ccaccgacac cgtcttccgc gtcaccgcca 4260
ccggggacgg cttcacctcg cgggaggtgc cgggtggacc gccctgacc aagggtgttc 4320
ccgacgacga atcggacgac gaatcggacg ccggggagga cggcgaccgc gactcccgga 4380
cgttcgtcgc gtacggg

```

<210> 49

<211> 367

<212> PRT

<213> Escherichia coli

<400> 49

```

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

```

Pro	Ala	Arg	Leu	Leu	Ala	Ala	Gly	Ile	Gly	Asp	Ala	Leu	Ala	Thr	Trp
			165						170					175	
Phe	Glu	Ala	Arg	Ala	Cys	Ser	Arg	Ser	Gly	Ala	Thr	Thr	Met	Ala	Gly
			180						185				190		
Gly	Lys	Cys	Thr	Gln	Ala	Ala	Leu	Ala	Leu	Ala	Glu	Leu	Cys	Tyr	Asn
		195					200					205			
Thr	Leu	Leu	Glu	Glu	Gly	Glu	Lys	Ala	Met	Leu	Ala	Ala	Glu	Gln	His
	210					215					220				
Val	Val	Thr	Pro	Ala	Leu	Glu	Arg	Val	Ile	Glu	Ala	Asn	Thr	Tyr	Leu
225					230					235					240
Ser	Gly	Val	Gly	Phe	Glu	Ser	Gly	Gly	Leu	Ala	Ala	Ala	His	Ala	Val
			245						250					255	
His	Asn	Gly	Leu	Thr	Ala	Ile	Pro	Asp	Ala	His	His	Tyr	Tyr	His	Gly
		260						265					270		
Glu	Lys	Val	Ala	Phe	Gly	Thr	Leu	Thr	Gln	Leu	Val	Leu	Glu	Asn	Ala
	275						280					285			
Pro	Val	Glu	Glu	Ile	Glu	Thr	Val	Ala	Ala	Leu	Ser	His	Ala	Val	Gly
	290					295					300				
Leu	Pro	Ile	Thr	Leu	Ala	Gln	Leu	Asp	Ile	Lys	Glu	Asp	Val	Pro	Ala
305					310					315					320
Lys	Met	Arg	Ile	Val	Ala	Glu	Ala	Ala	Cys	Ala	Glu	Gly	Glu	Thr	Ile
			325						330					335	
His	Asn	Met	Pro	Gly	Gly	Ala	Thr	Pro	Asp	Gln	Val	Tyr	Ala	Ala	Leu
		340						345					350		
Leu	Val	Ala	Asp	Gln	Tyr	Gly	Gln	Arg	Phe	Leu	Gln	Glu	Trp	Glu	
		355					360					365			

<210> 50

<211> 1120

<212> DNA

<213> Artificial Sequence

<220>

<223> nucleotide sequence encoding E.coli gldA codon-optimised for yeast

<400> 50

ctgcagaaaa	tgacagaat	catccaatct	ccaggtaagt	acatccaagg	tgctgacgtt	60
atcaacagat	tggtgaata	cttgaagcca	ttggctgaaa	gatgggttgg	tggttggtgac	120
aagttcggtt	tggttttcgc	tcaatctacc	gttgaaaagt	ctttcaagga	cgctgggttg	180
gttggtgaaa	tcgctccatt	cggtggtgaa	tggtctcaaa	acgaaatcga	cagattgaga	240
ggtatcgctg	aaaccgctca	atgtggtgct	atcttgggta	tcggtggtgg	taagaccttg	300
gacaccgcta	aggctttggc	tcacttcatg	ggtgttccag	ttgctatcgc	tccaaccatc	360
gcttctaccg	acgctccatg	ttctgctttg	tctgttatct	acaccgacga	aggtgaattc	420
gacagatact	tggtgttgcc	aaacaacca	aacatgggta	tcgttgacac	caagatcggt	480
gctggtgctc	cagctagatt	gttggcagct	ggtatcggtg	acgctttggc	tacctggttc	540
gaagctagag	cttggtctag	atctggtgct	accaccatgg	ctggtggtaa	gtgtacccaa	600
gctgctttgg	ctttggctga	attgtgttac	aacaccttgt	tggaagaagg	tgaaaaggct	660
atgttggtcg	ctgaacaaca	cggtgttacc	ccagctttgg	aaagagttat	cgaagctaac	720
acctacttgt	ctggtgttgg	tttogaatct	ggtggtttgg	ctgctgctca	cgctgttcac	780
aacggtttga	ccgctatccc	agacgctcac	cactactacc	acggtgaaaa	ggttgctttc	840
ggtaccttga	cccaattggg	tttggaatac	gtccagttg	aagaaatcga	aaccgttgct	900
gctttgtctc	acgctgttgg	tttgccaatc	accttggtc	aattggacat	caaggaagac	960
gttccagcta	agatgagaat	cggtgctgaa	gctgcttgtg	ctgaagggtga	aaccatccac	1020

aacatgccag gtggtgctac cccagaccaa gtttacgctg ctttggttggg tgctgaccaa 1080
tacgggtcaaa gattcctaca agaatgggaa taaggcgcg 1120

<210> 51
<211> 1949
<212> DNA
<213> Artificial Sequence

<220>
<223> pRNglA

<400> 51
actagtaa at gtgtggggaa gcgggtaagc tgccacagca attaatgcac aacatttfaac 60
ctacattctt ctttatcgga tcctcaaaac ccttaaaaac atatgcctca ccctaacata 120
ttttccaatt aaccctcaat atttctctgt caccggcct ctattttcca ttttcttctt 180
taccogccac gcgttttttt ctttcaaatt tttttcttcc ttcttctttt tottccacgt 240
cctcttgc at aaataaataa accgttttga aaccaaaact gcctctctct ctcctttttg 300
aaatatTTTT gggtttgttt gatcctttcc ttcccaatct ctcttgttta atatatattc 360
atttatatca cgctctcttt ttatcttctt ttttttcttc tctcttgat tcttccttcc 420
cctttctact caaaccaaga agaaaaagaa aagggtcaatc tttgttaaag aataggatct 480
tctactacat cagcttttag atttttcacg cttactgctt ttttcttccc aagatcgaaa 540
atttactgaa ttaactgcag aaaatggaca gaatcatcca atctccaggt aagtacatcc 600
aagggtgctga cgttatcaac agattgggtg aatacttgaa gccattggct gaaagatggg 660
tggttggttg tgacaagtgc gttttgggtt tcgctcaatc taccgttgaa aagtctttca 720
aggacgctgg tttggttggt gaaatcgctc catcggtggg tgaatgttct caaaacgaaa 780
tcgacagatt gagaggatc gctgaaaccg ctcaatgtgg tgctatcttg ggtatcggtg 840
gtggtaagac cttggacacc gctaaggctt tggtcactt catgggtgtt ccagttgcta 900
tcgctccaac catcgcttct accgacgctc catgttctgc tttgtctgtt atctacaccg 960
acgaagggtga attcgacaga tacttggtgt tgccaaacaa ccaaacatg gttatcggtg 1020
acaccaagat cgttgctggg gctccagcta gattggtggc agctggatc ggtgacgctt 1080
tggttacctg gttcgaagct agagcttggt ctagatctgg tgctaccacc atggctgggtg 1140
gtaagtgtac ccaagctgct ttggctttgg ctgaattgtg ttacaacacc ttgttggaag 1200
aagggtgaaa ggctatgttg gctgctgaac aacacgttgt taccctagct ttggaaagag 1260
ttatcgaagc taacacctac ttgtctgggt ttggtttcga atctgggtgg ttggctgctg 1320
ctcacgctgt tcacaacggt ttgacgcta tcccagacgc tcaccactac taccacggtg 1380
aaaagggtgc tttcggtacc ttgacccaat tggttttgga aaacgctcca gttgaagaaa 1440
tcgaaaccgt tgctgctttg tctcacgctg ttggtttgcc aatcaccttg gctcaattgg 1500
acatcaagga agacgttcca gctaagatga gaatcggtgc tgaagctgct tgtgctgaag 1560
gtgaaaccat ccacaacatg ccagggtggg ctaccccaga ccaagtttac gctgctttgt 1620
tggttgctga ccaatacggg caaagattcc tacaagaatg ggaataaggc gcgccccttt 1680
tcctttgtcg atatcatgta attagttatg tcacgcttac attcacgcc tectcccaca 1740
tccgctctaa ccgaaaagga aggagttaga caacctgaag tctaggtccc tatttatattt 1800
ttttaatagt tatgttagta ttaagaacgt tatttatatt tcaaattttt cttttttttc 1860
tgtacaaacg cgtgtacgca tgtaacatta tactgaaaac cttgcttgag aaggttttgg 1920
gacgctcgaa ggcttcctag gcgtacgtt 1949

<210> 52
<211> 552
<212> PRT
<213> *Citrobacter freundii*

<400> 52
Met Ser Gln Phe Phe Phe Asn Gln Arg Thr His Leu Val Ser Asp Val
1 5 10 15

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

Lys Phe Tyr Gly Gly Ala Asp Glu Gly Asp Arg Thr Met Ile Asp Ala
 465 470 475 480
 Leu Gln Pro Ala Leu Thr Ser Leu Leu Thr Gln Pro Gln Asn Leu Gln
 485 490 495
 Ala Ala Phe Asp Ala Ala Gln Ala Gly Ala Glu Arg Thr Cys Leu Ser
 500 505 510
 Ser Lys Ala Asn Ala Gly Arg Ala Ser Tyr Leu Ser Ser Glu Ser Leu
 515 520 525
 Leu Gly Asn Met Asp Pro Gly Ala His Ala Val Ala Met Val Phe Lys
 530 535 540
 Ala Leu Ala Glu Ser Glu Leu Gly
 545 550

<210> 53
 <211> 1678
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> nucleotide sequence encoding the *C. freundii* dhaK codon-optimised
 for yeast

<400> 53
 tctagaaaaa tgtctcaatt cttcttcaac cagagaaccc acttggtttc tgacgttata 60
 gacgggtgcta tcatcgcttc accatggaac aatttggtta gattggaata tgaccagat 120
 atcagaatcg ttgttagaag agacttgaac aagaacaacg ttgctgttat ctctggtggt 180
 ggttctggtc acgaaccagc tcacgttggg ttcatcggtta agggatatgtt gaccgctgct 240
 gtttgtggtg acgttttgcg ttctccatct gttgacgctg ttttgactgc tatccaagct 300
 gttaccggtg aagctggttg tttgttgatc gttaagaact acaccggtga cagattgaac 360
 ttcggttttg ctgctgaaaa ggctagaaga ttgggttaca acgttgaaat gttgatcggt 420
 ggtgacgaca tctctttgcc agacaacaag caccacaagag gtatcgctgg taccatcttg 480
 gttcacaaga tcgctggtta cttcgttgaa agaggttaca acttagctac cgttttgaga 540
 gaagctcaat acgctgcttc taacaccttc tctttgggtg ttgctttgtc ttcttgtcac 600
 ttgccacaag aaaccgacgc tgetccaaga caccacccag gtcacgctga attgggtatg 660
 ggtatccacg gtgaaccagg tgcttctggt atcgacaccc aaaactctgc tcaagttggt 720
 aacttgatgg ttgacaagtt gttggctgct ttgccagaaa ccggtagatt ggctgttatg 780
 atcaacaact tgggtggtgt ttctgttgct gaaatggcta tcatcaccag agaattggct 840
 tcttctccat tgcactcaag aatcgactgg ttgatcggtc cagcttcttt ggtaaccgct 900
 ttggacatga agggtttctc tttgaccgct atcgtttttg aagaatctat cgaaaaggct 960
 ttgttgaccg aagttgaaac ctctaactgg ccaaccccag ttccaccaag agaaatcacc 1020
 tgtgttggtt cttctcacgc ttctgctaga gttgaattcc aaccatctgc taacgctttg 1080
 gttgctggtg tcgttgaatt ggttaccgct acctgtctg acttggaac ccacttgaac 1140
 gctttggacg ctaagggttg tgacggtgac accggttcta ccttcgctgc tgotgctaga 1200
 gaaatcgctt ctttggtgca cagacaacaa ttgccattga acaacttggc taccttggtc 1260
 gctttgatcg gtgaaagatt gaccgttggt atgggtggtt cttctggtgt tttgatgtct 1320
 atcttcttca ccgctgctgg tcaaaagttg gaacaagggt ctaacgttgt tgaagctttg 1380
 aacaccggtt tggtcaaat gaagttctac ggtggtgctg acgaagggtga cagaaccatg 1440
 atcgacgctt tgcaaccagc ttgacctct ttggtggctc aaccaaagaa cttgcaagct 1500
 gctttcgacg ctgctcaagc tgggtgctgaa agaacctgtt tgtcttctaa ggetaacgct 1560
 ggtagagctt ctacttgct ttctgaatct ttggtgggtg acatggaccc aggtgctcaa 1620
 agattggcta tggttttcaa ggctttggct gaatctgaat tgggttaata aggtcgac 1678

<210> 54
 <211> 2951

<212> DNA

<213> Artificial Sequence

<220>

<223> pRNdhak

<400> 54

ggatccacta	gtaacggccg	ccagtgtgct	ggaattcgcc	cttctcgagc	ttaagacgcg	60
tttcttcttc	agattccctc	atggagaaaag	tgcggcagat	gtatatgaca	gagtcgccag	120
tttccaagag	actttattca	ggcacttcca	tgataggcaa	gagagaagac	ccagagatgt	180
tggtgtccta	gttacacatg	gtattttattc	cagagtattc	ctgatgaaat	ggtttagatg	240
gacatacga	gagtttgaat	cgtttacca	tggtccctaac	gggagcgtaa	tggtgatgga	300
actggacga	tccatcaata	gatacgtcct	gaggaccgtg	ctacccaaat	ggactgattg	360
tgagggagac	ctaactacat	agtgtttaaa	gattacggat	atttaactta	cttagaataa	420
tgccattttt	ttgagttata	ataatcctac	gttagtgtga	gcgggattta	aactgtgagg	480
accttaatac	attcagacac	ttctgcggta	tcaccctact	tattcccttc	gagattatat	540
ctaggaaccc	atcaggttgg	tggaagatta	cccggttctaa	gacttttcag	cttcctctat	600
tgatgttaca	cctggacacc	ccttttctgg	catccagttt	ttaatcttca	gtggcatgtg	660
agattctccg	aaattaatta	aagcaatcac	acaattctct	cggataccac	ctcggttgaa	720
actgacaggt	ggtttggttac	gcatgcta	gcaaaggagc	ctatatacct	ttggctcggc	780
tgctgtaaca	gggaatataa	agggcagcat	aatttaggag	tttagtgaa	ttgcaacatt	840
tactattttc	ccttcttaacg	taaatatttt	tctttttaat	tctaaatcaa	tctttttcaa	900
ttttttgttt	gtattctttt	cttgcttaaa	tctataacta	caaaaaacac	atacataaat	960
ctagaaaaat	gtctcaattc	ttcttcaacc	agagaaccca	cttggtttct	gacgttatcg	1020
acggtgctat	catcgcttca	ccatggaaca	atttggttag	attggaatct	gaccagcta	1080
tcagaatcgt	tgtagaaga	gacttgaaca	agaacaacgt	tgctgttatc	tctggtggtg	1140
gttctggta	cgaaccagct	cacgttggtt	tcacggtaa	gggtatgttg	accgctgctg	1200
tttgtggtga	cgttttcgtc	tctccatctg	ttgacgctgt	tttgactgct	atccaagctg	1260
ttaccgggtga	agctggttgt	ttggtgatcg	ttaagaacta	caccgggtgac	agattgaact	1320
tcggtttggc	tgctgaaaag	gctagaagat	tggtgttaca	cggtgaaatg	ttgatcggtg	1380
gtgacgacat	ctctttgcca	gacaacaagc	accaagagg	tatcgctggt	accatcttgg	1440
ttcacaagat	cgctgggttac	ttcgctgaaa	gaggttaca	cttagctacc	gttttgagag	1500
aagctcaata	cgctgcttct	aacaccttct	ctttgggtgt	tgctttgtct	tcttgctact	1560
tgccacaaga	aaccgacgct	gctccaagac	accacccagg	tcacgctgaa	ttgggtatgg	1620
gtatccacgg	tgaaccaggt	gcttctgtta	tcgacacca	aaactctgct	caagttgtta	1680
acttgatggt	tgacaagttg	ttggctgctt	tgccagaaac	cggtagattg	gctgttatga	1740
tcaacaactt	gggtggtggt	tctgttgctg	aaatggctat	catcaccaga	gaattggctt	1800
cttctccatt	gcaactcaaga	atcgactggt	tgatcggtcc	agcttctttg	gtaaccgctt	1860
tggaacatgaa	gggtttctct	ttgaccgcta	tcgttttgga	agaatctatc	gaaaaggctt	1920
tggtgaccga	agttgaaacc	tctaactggc	caaccccgct	tccaccaaga	gaaatcacct	1980
gtgttggttc	ttctcacgct	tctgctagag	ttgaattcca	accatctgct	aacgctttgg	2040
ttgctggtat	cgttgaattg	gttaccgcta	ccttgctctga	cttggaacc	cacttgaacg	2100
ctttggacgc	taagggttgg	gacggtgaca	ccggttctac	cttcgctgct	gctgctagag	2160
aaatcgcttc	tttggttcac	agacaacaat	tgccattgaa	caacttggct	acottgttctg	2220
ctttgatcgg	tgaaagattg	accgttggtta	tggttggttc	ttctgggtgtt	ttgatgtcta	2280
tcttcttcac	cgctgctggg	caaaagttgg	aacaagggtgc	taacgttggt	gaagctttga	2340
acaccggttt	ggctcaa	aagttctacg	gtggtgctga	cgaagggtgac	agaaccatga	2400
tcgacgcttt	gcaaccagct	ttgacctctt	tggttggtca	accaaagaac	ttgcaagctg	2460
ctttcgacgc	tgctcaagct	ggtgctgaaa	gaacctgttt	gtcttctaag	gctaaccgctg	2520
gtagagcttc	ttacttgtct	tctgaatctt	tggtgggttaa	catggacce	ggtgctcaaa	2580
gattggctat	ggttttcaag	gctttggctg	aatctgaatt	gggttaataa	ggtcgagaca	2640
aatcgctctt	aaatatatac	ctaaagaaca	ttaaagctat	attataagca	aagatacgt	2700
aattttgctt	atattattat	acacatatca	tatttctata	tttttaagat	ttggttatat	2760
aatgtacgta	atgcaaagga	aataaatttt	atacattatt	gaacagcgtc	caagtaacta	2820
cattatgtgc	actaatagtt	tagcgtcgtg	aagactttat	tgtgtcgcga	aaagtaaaaa	2880

ttttaaaaat tagagcacct tgaacttgcg aaaaagggttc tcatcaactg tttaaaacgt 2940
acgaagctta a 2951

<210> 55
<211> 14200
<212> DNA
<213> Artificial Sequence

<220>
<223> pRN957

<400> 55
gacgaaaggg cctcgtgata cgctatTTTT tataggTTaa tgtcatgata ataagacgtc 60
aggtggcact tttcggggaa atgtgcgcgg aacccttatt tgTTTTatTT tctaaataca 120
ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 180
aaggaagagt atgagtattc aacatttccg tgcgcgccct attccctttt ttgcggcatt 240
ttgccttctt gtttttgcgc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 300
gttggttgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 360
ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 420
ggtattatcc cgtattgacg ccgggcaaga gcaactcggc cgcgcatac actattctca 480
gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 540
aagagaatta tgcagtgtcg ccataaccat gagtgataac actgcggcca acttacttct 600
gacaacgacg ggaggaccga aggagctaac cgctTTTTtt cacaacatgg gggatcatgt 660
aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 720
caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 780
tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 840
acttctgcgc tcggcccttc cggtggctg gtttattgct gataaatctg gagccggtga 900
gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 960
agttatctac acgacgggca gtcaggcaac tatggatgaa cgaaatagac agatcgctga 1020
gataggtgcc tcaactgatta agcattggta actgtcagac caagtttact catatatact 1080
ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tcctttttga 1140
taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccg 1200
agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 1260
aacaacaaaa ccaccgctac cagcgggtgg ttgtttgccg gatcaagagc taccaactct 1320
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgt 1380
gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 1440
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggaactc 1500
aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggt cgtgcacaca 1560
gccagccttg gagcgaacga cctacaccga actgagatac ctacagcgtg agcattgaga 1620
aagcgcacg ctcccggaag ggagaaaggc ggacaggat cgggtaagcg gcagggtcgg 1680
aacaggagag cgcacgaggg agcttccagg ggggaacgcc tggatatctt atagtcctgt 1740
cgggtttcgc cactctgac ttgagcgtcg atttttgtga tgctcgtcag gggggccgag 1800
cctatggaaa aacgccagca acgcggcctt ttacgggttc ctggcctttt gctggccttt 1860
tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 1920
tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcagat cagtgagcga 1980
ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta 2040
atgcagctgg cacgacaggt tccccgactg gaaagcgggc agtgagcgca acgcaattaa 2100
tgtgagttac ctcactcatt aggcacccca ggctttacac ttatgcttc cggctcctat 2160
gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta 2220
cgccaagctc ggaattaacc ctcaactaaag ggaacaaaag ctgggtaccg ggccccccct 2280
cgaggctcag cttaagacgc gtttcttctt cagattccct catggagaaa gtgcggcaga 2340
tgtatatgac agagtcgcca gtttccaaga gactttatcc aggcacttcc atgataggca 2400
agagagaaga cccagagatg ttgttgctct agttacacat ggtatttatt ccagagtatt 2460
cctgatgaaa tggtttagat ggacatacga agagtttgaa tcgtttacca atgttcctaa 2520
cgggagcgtg atggtgatgg aactggacga atccatcaat agatacgtcc tgaggaccgt 2580

gctacccaaa	tggactgatt	gtgagggaga	cctaactaca	tagtgtttaa	agattacgga	2640
tatttaactt	acttagaata	atgccatfff	tttgagttat	aataatccta	cgttagtgtg	2700
agcgggattt	aaactgtgag	gaccttaata	cattcagaca	cttctgcggt	atcaccctac	2760
ttattccctt	cgagattata	tctaggaacc	catcaggttg	gtggaagatt	acccgttcta	2820
agacttttca	gcttcctcta	ttgatgttac	acctggacac	cccttttctg	gcatccagtt	2880
tttaatcttc	agtggcatgt	gagattctcc	gaaattaatt	aaagcaatca	cacaattctc	2940
tgggatacca	cctcggttga	aactgacagg	tggtttggtta	cgcatgctaa	tgcaaaggag	3000
cctatatacc	tttggctcgg	ctgctgtaac	agggaatata	aagggcagca	taatttagga	3060
gtttagtga	cttgcaacat	ttactatfff	cccttcctac	gtaaatatff	ttctttttaa	3120
ttctaaatca	atctttttca	atftttttgtt	tgtattcttt	tcttgcttaa	atctataact	3180
acaaaaaaca	catacataaa	tctagaaaaa	tgtctcaatt	cttcttcaac	cagagaaccc	3240
acttggtttc	tgacgttata	gacggtgcta	tcatcgcttc	accatggaac	aatttggtta	3300
gattggaatc	tgacccagct	atcagaatcg	ttgttagaag	agacttgaac	aagaacaacg	3360
ttgctgttat	ctctggtggt	ggttctggtc	acgaaccagc	tcacgttggt	ttcatcggtta	3420
agggtatggt	gaccgctgct	gtttgtgggtg	acgttttctgc	ttctccatct	gttgacgctg	3480
ttttgactgc	tatccaagct	gttaccgggtg	aagctggttg	tttggtgatc	gttaagaact	3540
acaccggtga	cagattgaac	ttcggttttg	ctgctgaaaa	ggctagaaga	ttgggttaca	3600
acgttgaaat	gttgatcggt	ggtgacgaca	tctctttgcc	agacaacaag	cacccaagag	3660
gtatcgctgg	taccatcttg	gttcacaaga	tctgctggtta	cttcgctgaa	agaggttaca	3720
acttagctac	cgtttttgaga	gaagctcaat	acgctgcttc	taacaccttc	tctttgggtg	3780
ttgctttgtc	ttcttgtcac	ttgccacaag	aaaccgacgc	tgctccaaga	caccaccag	3840
gtcacgctga	attgggtatg	ggtatccacg	gtgaaccagg	tgcttctggt	atcgacaccc	3900
aaaactctgc	tcaagttggt	aacttgatgg	ttgacaagtt	gttggtctgt	ttgccagaaa	3960
ccggtagatt	ggctgttatg	atcaacaact	tgggtggtgt	ttctgttgct	gaaatggcta	4020
tcatcaccag	agaattggct	tcttctccat	tgcactcaag	aatcgactgg	ttgatcggtc	4080
cagcttcttt	ggtaaccgct	ttggacatga	agggtttctc	tttgaccgct	atcgtttttg	4140
aagaatctat	cgaaaaggct	ttggtgaccg	aagttgaaac	ctctaactgg	ccaaccccag	4200
ttccaccaag	agaaatcacc	tgtgttggtt	cttctcacgc	ttctgctaga	gttgaattcc	4260
aaccatctgc	taacgctttg	gttgctggta	tctgtgaatt	ggttaccgct	accttgtctg	4320
acttggaac	ccacttgaac	gctttggacg	ctaagggttg	tgacggtgac	accggttcta	4380
ccttcgctgc	tgctgctaga	gaaatcgctt	ctttgttgca	cagacaacaa	ttgccattga	4440
acaacttggc	taccttggtc	gctttgatcg	gtgaaagatt	gaccgttggt	atgggtgggt	4500
cttctggtgt	tttgatgtct	atcttcttca	ccgctgctgg	tcaaaagttg	gaacaagggtg	4560
ctaacgttgt	tgaagctttg	aacaccgggt	tggctcaaat	gaagttctac	ggtggtgctg	4620
acgaagggtga	cagaaccatg	atcgacgctt	tgcaaccagc	tttgacctct	ttgttggtct	4680
aaccaaaaga	cttgcaagct	gctttcgacg	ctgctcaagc	tggtgctgaa	agaacctggt	4740
tgtcttctaa	ggctaaccgt	ggtagagctt	cttacttgct	ttctgaatct	ttgttgggta	4800
acatggaccc	aggtgctcaa	agattggcta	tggttttcaa	ggctttggct	gaatctgaat	4860
tgggttaata	aggtcgagac	aaatcgctct	taaatatata	cctaaagaac	attaaagcta	4920
tattataagc	aaagatacgt	aaatftttgct	tatattatta	tacacatata	atatttctat	4980
atfttttaaga	tttggttata	taatgtacgt	aatgcaaagg	aaataaatft	tatacatftat	5040
tgaacagcgt	ccaagtaact	acattatgtg	cactaatagt	ttagcgtcgt	gaagactfta	5100
ttgtgtcgcg	aaaagtaaaa	atftttaaaaa	ttagagcacc	ttgaacttgc	gaaaaagggt	5160
ctcatcaact	gttttaaaacg	tacgcctagg	aagccttctga	gcgtcccaaa	accttctcaa	5220
gcaaggfttt	cagtataatg	ttacatgcgt	acacgcgftt	gtacagaaaa	aaaagaaaaa	5280
tttgaaatat	aaataacggt	cttaataacta	acataactat	taaaaaaaat	aaatagggac	5340
ctagacttca	ggttgtctaa	ctccttctct	ttcggttaga	gcggatgtgg	gaggagggcg	5400
tgaatgtaag	cgtgacataa	ctaattacat	gatatcgaca	aaggaaaagg	ggcgcgcctt	5460
attcccattc	ttgtagggaat	ctttgaccgt	attggtcagc	aaccaacaaa	gcagcgtaaa	5520
cttggtctgg	ggttagcacca	cctggcatgt	tgtggatggt	ttcaccttca	gcacaagcag	5580
cttcagcaac	gattctcatc	ttagctggaa	ogtcttctct	gatgtccaat	tgagccaagg	5640
tgattggcaa	accaacagcg	tgagacaaag	cagcaacggt	ttcgatttct	tcaactggag	5700
cgttttccaa	aaccaattgg	gtcaagggtac	cgaaagcaac	cttttcaccg	tggtagtagt	5760
ggtgagcgtc	tgggatagcg	gtcaaaccgt	tgtgaacagc	gtgagcagca	gccaaccac	5820
cagattcgaa	accaacacca	gacaagtagg	tgttagcttc	gataactctt	tccaaagctg	5880
gggtaacaac	gtgttggttca	gcagccaaca	tagccttttc	accttcttcc	aacaagggtg	5940

tgtaacacaa	ttcagccaaa	gccaaagcag	cttgggtaca	cttaccacca	gccatggtgg	6000
tagcaccaga	tctagaacaa	gctctagctt	cgaaccaggt	agccaaagcg	tcaccgatac	6060
cagctgccaa	caatctagct	ggagcaccag	caacgatctt	ggtgtcaacg	ataaccatgt	6120
ttgggttggt	tggaacaac	aagtatctgt	cgaattcacc	ttcgtcgggtg	tagataacag	6180
acaaagcaga	acatggagcg	tcggtagaag	cgatgggttg	agcgatagca	actggaacac	6240
ccatgaagtg	agccaaagcc	ttagcgggtg	ccaaggtctt	accaccaccg	ataccaaga	6300
tagcaccaca	ttgagcgggt	tcagcgatac	ctctcaatct	gtcgatttctg	ttttgagaac	6360
attcaccacc	gaatggagcg	atttcaacaa	ccaaaccagc	gtccttgaaa	gacttttcaa	6420
cggtagattg	agcgaaaccc	aaaacgaact	tgtcaccaac	aaccaaccat	ctttcagcca	6480
atggcttcaa	gtattcacc	aatctgttga	taacgtcagc	accttggtatg	tacttacctg	6540
gagattggat	gattctgtcc	attttctgca	gttaattcag	taaattttctg	atcttgggaa	6600
gaaaaaagca	gtaagcgtga	aaaatctaaa	agctgatgta	gtagaagatc	ctattcttta	6660
acaaagattg	accttttctt	tttcttcttg	gtttgagtag	aaaggggaag	gaagaataca	6720
agagagagga	aaaaaaggaa	gataaaaaga	gagcgtgata	taaatgaata	tatattaac	6780
aagagagatt	gggaaggaaa	ggatcaaaca	aacccaaaaa	tatttcaaaa	aggagagaga	6840
gaggcgagtt	tggtttcaaa	acggtttatt	tatttatgca	agaggacgtg	gaagaaaaag	6900
aagaaggaag	aaaaaaattt	gaaagaaaaa	aacgcgtggc	gggtaaagaa	gaaaatggaa	6960
aatagaggcc	gggtgacaga	gaaatattga	gggttaattg	gaaaatatgt	taggggtgagg	7020
catatgtttt	taagggtttt	gaggatccga	taaggaagaa	tgtaggttaa	atgttgtgca	7080
ttaattgctg	tggcagctta	cccgcttccc	cacacattta	ctagtctcga	gctcttcaac	7140
tcaagacgca	cagatattat	aacatctgca	taataggcat	ttgcaagaat	tactcgtgag	7200
taaggaaaga	gtgaggaact	atcgcatacc	tgcattttaa	gatgccgatt	tgggcgcgaa	7260
tcctttatatt	tggtttcacc	ctcatactat	tatcagggcc	agaaaaagga	agtgtttccc	7320
tccttcttga	attgatgtta	ccctcataaa	gcacgtggcc	tcttatcgag	aaagaaatta	7380
ccgtcgtctg	tgatttggtt	gcaaaaagaa	caaaactgaa	aaaaccagaa	cacgctcgac	7440
ttcctgtctt	cctattgatt	gcagcttcca	atttcgtcac	acaacaaggt	cctagcgacg	7500
gctcacaggt	tttgtaacaa	gcaatcgaag	gttctggaat	ggcgggaaag	ggttttagtac	7560
cacatgctat	gatgcccaact	gtgatctcca	gagcaaagtt	cgttcgatcg	tactgttact	7620
ctctctcttt	caaacagaat	tgtccgaatc	gtgtgacaac	aacagcctgt	tctcacacac	7680
tcttttcttc	taaccaaggg	ggtggttttag	tttagtagaa	cctcgtgaaa	cttacattta	7740
catatatata	aacttgcata	aattggtcaa	tgcaagaaat	acatatattg	tcttttctaa	7800
ttcgtagttt	ttcaagttct	tagatgcttt	ctttttctct	tttttacaga	tcatacagga	7860
agtaattatc	tactttttac	aacaaatata	tctagaaaat	ggctgttacc	aacgttgctg	7920
aattgaacgc	tttggttgaa	agggttaaga	aggtcacaag	agaatacgtc	tctttcaccc	7980
aagaacaagt	tgacaagatc	ttcagagctg	ctgctttggc	tgtcgtgac	gctagaatcc	8040
cattggctaa	gatggctggt	gctgaatctg	gtatgggtat	cgttgaagac	aaggttatca	8100
agaaccactt	cgttcttgaa	tacatctaca	acgtttacaa	ggacgaaaag	acctgtggtg	8160
ttttgtcaga	agacgacacc	ttcgggtacca	tcaccatcgc	tgaaccaatc	ggtatcatct	8220
gtgggtatcgt	tccaaccacc	aacccaacct	ctaccgctat	cttcaagtct	ttgatctctt	8280
tgaagaccag	aaacgctatc	atcttctctc	cacaccaag	agctaaagac	gctaccaaca	8340
aggctgctga	catcgttttg	caagctgcta	tcgctgctgg	tgctccaaag	gacttgatcg	8400
gttggtatcga	ccaaccatct	gttgaaattgt	ctaacgcttt	gatgcaccac	ccagacatca	8460
acttgatctt	ggctaccggt	ggtccaggta	tggttaaggc	tgttactct	tctggtaagc	8520
cagctatcgg	tgttggtgct	ggtaacaccc	cagttgttat	cgacgaaacc	gctgacatca	8580
agagagctgt	tgcttctggt	ttgatgtcta	agaccttcga	caacgggtgtt	atctgtgctt	8640
ctgaacaatc	tgttggtggt	gttgactctg	tttacgacgc	tgtagagaa	agattcgcta	8700
cccacgggtg	ttacttggtg	caaggtaagg	aattgaaggc	tgttcaagac	gttatcttga	8760
agaacgggtgc	tttgaacgct	gctatcggtg	gtcaaccagc	ttacaagatc	gctgaattag	8820
ctggtttctc	tgttccagaa	aacaccaaga	tcttgatcgg	tgaagttacc	gttggtgacg	8880
aatctgaacc	attcgctcac	gaaaagttgt	ctccaacctt	ggctatgtac	agagctaagg	8940
acttcgaaga	cgtgttgtaa	aaagctgaaa	agttgggttg	tatgggtggg	attggtcaca	9000
cctcttggtt	gtacaccgac	caagacaacc	aaccagctag	agtttcttac	ttcgggtcaaa	9060
agatgaagac	cgtagaatc	ttgatcaaca	cccagcttc	tcaaggtggg	atcggtgact	9120
tgtacaactt	caagttggct	ccatctttga	ccttgggttg	tggttcttgg	ggtggtaact	9180
ctatctctga	aaacggtggg	ccaaagcact	tgatcaacaa	gaagaccggt	gctaagagag	9240
ctgaaaacat	gttgtggcac	aagttgccaa	aatctatcta	cttcagaaga	ggttctttgc	9300

caatcgcttt	ggaacgaagt	atcaccgacg	gtcacaagag	agctttgatc	gttaccgaca	9360
gattcttggt	caacaacggg	tacgctgacc	aaatcacctc	tgttttgaag	gctgctgggt	9420
ttgaaaccga	agttttcttc	gaagttgaag	ctgacccaac	cttgtctatc	gttagaaagg	9480
gtgctgaatt	ggctaactct	ttcaagccag	acgttatcat	cgctttgggt	ggtaggttct	9540
caatggacgc	tgctaagatc	atgtgggtta	tgtacgaaca	cccagaaacc	cacttcgaag	9600
aattggcttt	gagattcatg	gacatcagaa	agagaatcta	caagttccca	aagatgggtg	9660
ttaaggctaa	gatgatcgct	gttaccacca	cctctgggtac	cggttctgaa	gttaccctat	9720
tcgctgttgt	taccgacgac	gctaccgggc	aaaagtaccc	attggctgac	tacgctttga	9780
ccccagacat	ggctatcggt	gacgctaact	tggttatgga	catgccaaag	tctttgtgtg	9840
ctttcggtgg	tttggaacgt	gttaccacac	ctatggaagc	ttacgtttct	gttttggtct	9900
ctgaattctc	tgacgggtcaa	gctttgcaag	ctttgaagtt	gttgaaggaa	tacttgccag	9960
cttcttacca	cgaagggtct	aagaaccacg	ttgctagaga	aagagttcac	tctgctgcta	10020
ccatcgctgg	tatcgctttc	gctaaccgct	tcttgggtgt	ttgtcactct	atggctcaca	10080
agttgggttc	tcaattccac	atcccacacg	gtttggctaa	cgctttgttg	atctgtaacg	10140
ttatcagata	caacgctaac	gacaacccaa	ccaagcaaac	cgctttctct	caatacgaca	10200
gaccacaagc	tagaagaaga	tacgctgaaa	tcgctgacca	cttgggtttg	tctgctccag	10260
gtgacagaac	cgctgcaaag	atcgaaaagt	tggtggcttg	gttggaaacc	ttgaaggctg	10320
aattgggtat	cccaaagtct	atcagagaag	ctggtgttca	agaagctgac	ttcttggcta	10380
acgttgacaa	gttgtctgaa	gacgctttcg	acgaccaatg	taccggtgct	aaccaagat	10440
accattgat	ctctgaattg	aagcaaactct	tggtggacac	ctactacggg	agagactacg	10500
ttgaagggtga	aaccgctgct	aagaagggaag	ctgctccagc	taaggctgaa	aagaaggcta	10560
agaagtctgc	ttagcttaag	cgcgcgaatt	tcttatgatt	tatgattttt	attattaaat	10620
aagttataaa	aaaaataagt	gtatacaaat	tttaaagtga	ctcttagggt	ttaaaacgaa	10680
aattcttatt	cttgagtaac	tctttctctg	aggtcagggt	gctttctcag	gtatagcatg	10740
aggtegtct	tattgaccac	acctctaccg	gcatgccgag	caaatgcctg	caaatcgctc	10800
ccattttcac	ccaattgtag	atatgctaac	tccagcaatg	agttgatgaa	tctcggtgtg	10860
tattttatgt	cctcagagga	caacacctgt	tgtaatcggt	cttccacacg	tacgaagctt	10920
aaaagggcga	attctgcaga	tatccatcac	actggcggcc	gcatgctagc	tccggattat	10980
cgatgataag	ctgtcaaaga	tgagaattaa	ttccacggac	tatagactat	actagatact	11040
ccgtctactg	tacgatacac	ttccgctcag	gtccttgccc	tttaacgagg	ccttaccact	11100
cttttgttac	tctattgatc	cagctcagca	aaggcagtgt	gatctaagat	tctatcttcg	11160
cgatgtagta	aaactagcta	gaccgagaaa	gagactagaa	atgcaaaaag	cacttctaca	11220
atggctgcc	tcattattat	ccgatgtgac	gctgcagctt	ctcaatgata	ttcgaatacg	11280
ctttgaggag	atacagccta	atatccgaca	aactgtttta	cagatttacg	atcgtaactg	11340
ttaccctatc	ttgaattttg	aacatccgaa	cctgggaggt	ttccctgaaa	cagatagtat	11400
atttgaacct	gtataataat	atatagtcta	gcgctttacg	gaagacaatg	tatgtatttc	11460
ggttctctga	gaaactattg	catctattgc	ataggtaatc	ttgcacgtcg	catccccggg	11520
tcattttctg	cgtttccatc	ttgcacttca	atagcatatc	tttgttaacg	aagcatctgt	11580
gcttcatttt	gtagaacaaa	aatgcaacgc	gagagcgcta	atttttcaaa	caaagaatct	11640
gagctgcatt	tttacagaac	agaaatgcaa	cgcgaaagcg	ctattttacc	aacgaagaat	11700
ctgtgcttca	tttttgtaaa	acaaaaatgc	aacgcgacga	gagcgcta	ttttcaaaca	11760
aagaatctga	gctgcatttt	tacagaacag	aaatgcaacg	cgagagcgct	atttttacca	11820
caaagaatct	atacttcttt	tttgttctac	aaaaatgcat	cccagagcg	ctatttttct	11880
aacaaagcat	cttagattac	tttttttctc	ctttgtgcgc	tctataatgc	agtctcttga	11940
taactttttg	cactgtagggt	ccgttaagggt	tagaagaagg	ctactttggg	gtctattttc	12000
tcttccataa	aaaaagcctg	actccacttc	ccgcgtttac	tgattactag	cgaagctgcg	12060
gggtgcatttt	ttcaagataa	aggcatcccc	gattatattc	tataccgatg	tggattgcgc	12120
atactttgtg	aacagaaagt	gatagcggtg	atgattcttc	attgggtcaga	aaattatgaa	12180
cggtttcttc	tattttgtct	ctatatacta	cgtataggaa	atgtttacat	tttcgtattg	12240
ttttcgatcc	actctatgaa	tagttcttac	tacaattttt	ttgtctaaag	agtaatacta	12300
gagataaaca	taaaaaatgt	agaggctcgag	tttagatgca	agttcaagga	gcgaaagggt	12360
gatgggtagg	ttatataggg	atatagcaca	gagatatata	gcaaagagat	acttttgagc	12420
aatgtttgtg	gaagcgggat	tcgcaatgcc	ggctttcccc	gtcaagctct	aaatcggggg	12480
ctccctttag	ggttccgatt	tagtgcttta	cggcacctcg	accccaaaaa	acttgattag	12540
gggtgatggt	cacgtagtgg	gccatcgccc	tgatagacgg	tttttcgccc	tttgacgttg	12600
gagtcacagt	tctttaatag	tggactcttg	ttccaaactg	gaacaacact	caaccctatc	12660

tccgtctatt	cttttgattt	ataagggatt	ttgccgattt	cggcctattg	gttaaaaaat	12720
gagctgattt	aacaaaaatt	taacgcgaat	tttaacaaaa	tattaacgtt	tacaatttcc	12780
tgatgcggta	ttttctcctt	acgcctctgt	gcggtatatt	acaccgcata	tgatccgtcg	12840
agttcaagag	aaaaaaaaag	aaaaagcaaa	aagaaaaaag	gaaagcgcgc	ctcgttcaga	12900
atgacacgta	tagaatgatg	cattaccttg	tcctcttcag	tatcatactg	ttcgtataca	12960
tacttactga	cattcatagg	tatacatata	tacacatgta	tatatatcgt	atgctgcagc	13020
tttaaataat	cgggtgtcact	acataagaac	acottttggtg	gagggaacat	cgttggtacc	13080
attggggcag	gtggcttctc	ttatggcaac	cgcaagagcc	ttgaacgcac	tctcactacg	13140
gtgatgatca	ttcttgcttc	gcagacaatc	aacgtggagg	gtaattctgc	tagcctctgc	13200
aaagctttca	agaaaatgcg	ggatcatctc	gcaagagaga	tctcctactt	tctccctttg	13260
caaaccaagt	tcgacaactg	cgtacggcct	gttcgaaaga	tctaccaccg	ctctggaaag	13320
tgccctcatc	aaaggcgcga	atcctgatcc	aaaccttttt	actccacgcg	ccagtagggc	13380
ctcttttaaaa	gcttgaccga	gagcaatccc	gcagtcttca	gtgggtgtgat	ggctcgtctat	13440
gtgtaagtca	ccaatgcact	caacgattag	cgaccagccg	gaatgcttgg	ccagagcatg	13500
tatcatatgg	tccagaaacc	ctatacctgt	gtggacgtta	atcacttgcg	attgtgtggc	13560
ctgttctgct	actgcttctg	cctctttttc	tgggaagatc	gagtgtctta	tcgctagggg	13620
accacccttt	aaagagatcg	caatctgaat	cttggtttca	tttgtaatac	gctttactag	13680
ggctttctgc	tctgtcatct	ttgccttcgt	ttatcttgcc	tgctcatttt	ttagtatatt	13740
cttcgaagaa	atcacattac	tttatataat	gtataattca	ttatgtgata	atgccaatcg	13800
ctaagaaaaa	aaaagagtca	tccgctaggt	ggaaaaaaaa	aaatgaaaat	cattaccgag	13860
gcataaaaaa	atatagagtg	tactagagga	ggccaagagt	aatagaaaaa	gaaaattgcg	13920
ggaaaggact	gtgttatgac	ttccctgact	aatgcctgtg	tcaaacgata	cctggcagtg	13980
actcctagcg	ctcaccaagc	tcttaaaacg	gaattatggt	gcactctcag	tacaatctgc	14040
tctgatgccg	catagttaag	ccagccccga	caccgcgcaa	caccgcgtga	cgcgcctga	14100
cgggcttgct	tgctcccggc	atccgcttac	agacaagctg	tgaccgtctc	cgggagctgc	14160
atgtgtcaga	ggttttcacc	gtcatcaccg	aaacgcgcga			14200

<210> 56

<211> 11289

<212> DNA

<213> Artificial Sequence

<220>

<223> pRN958

<400> 56

gacgaaaggg	cctcgtgata	cgcctatttt	tataggttaa	tgtcatgata	ataagacgtc	60
aggtggcact	tttcggggaa	atgtgcgcgg	aaccocctatt	tgtttatatt	tctaaataca	120
ttcaaataat	tatccgctca	tgagacaata	accctgataa	atgcttcaat	aatattgaaa	180
aaggaagagt	atgagtattc	aacatttccg	tgctgcctct	attccctttt	ttgcggcatt	240
ttgccttctc	gtttttgctc	accagaaaac	gctgggtgaa	gtaaaagatg	ctgaagatca	300
gttgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	360
ttttcgcctc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	420
ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggg	cgcgcgatac	actattctca	480
gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	540
aagagaatta	tgacgtgctg	ccataaccat	gagtgataac	actgoggcca	acttacttct	600
gacaacgac	ggaggaccga	aggagctaac	cgcttttttt	cacaacatgg	gggatcatgt	660
aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	720
caccacgatg	cctgtagcaa	tggcaacaac	gttgccgcaa	ctattaactg	gcgaactact	780
tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	840
acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccggtga	900
gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt	960
agttatctac	acgacgggca	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	1020
gatagggtgc	tactgatta	agcattggta	actgtcagac	caagttttact	catatatact	1080
ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	1140

taatctcatg	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccgt	1200
agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	gctgcttgca	1260
aacaaaaaaa	ccaccgctac	cagcggtggt	ttgtttgccc	gatcaagagc	taccaactct	1320
ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgtg	1380
gocgtagtta	ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	tcgctctgct	1440
aatcctgtta	ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggaactc	1500
aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggt	cgtgcacaca	1560
gcccagcttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	agcattgaga	1620
aagcgcacg	cttcccgaag	ggagaaaggc	ggacagggtat	ccggtaagcg	gcagggtcgg	1680
aacaggagag	cgcacgaggg	agcttccagg	ggggaacgcc	tggtatcttt	atagtcctgt	1740
cgggtttcgc	cacctctgac	ttgagcgtcg	atTTTTgtga	tgtcgtcag	gggggccgag	1800
cctatggaaa	aacgccagca	acgcggcctt	tttacgggttc	ctggcctttt	gctggccttt	1860
tgtcacatg	ttcttttctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	1920
tgagtgagct	gataccgctc	gccgcagccg	aacgaccgag	cgcagcaggt	cagtgagega	1980
ggaagcggaa	gagcgcacca	tacgcaaacc	gcctctcccc	gcgcgttgge	cgattcatta	2040
atgcagctgg	cacgacaggt	ttcccgaactg	gaaagcgggc	agtgagecga	acgcaattaa	2100
tgtgagttac	ctcactcatt	aggcacccca	ggctttacac	tttatgcttc	cggctcctat	2160
gttgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	2220
cgccaagctc	ggaattaacc	ctcactaaag	ggaacaaaag	ctgggtaccg	ggccctagga	2280
agccttcgag	cgtcccaaaa	ccttctcaag	caaggttttc	agtataatgt	tacatgcgta	2340
cacgcgtttg	tacagaaaaa	aaagaaaaat	ttgaaatata	aataacgttc	ttaataactaa	2400
cataactatt	aaaaaaaaata	aatagggacc	tagacttcag	gttgtctaac	tccttccttt	2460
tcggttagag	cggatgtggg	aggagggcgt	gaatgtaagc	gtgacataac	taattacatg	2520
atatcgacaa	aggaaaaggg	gcgcgcctta	ttcccattct	tgtaggaatc	tttgaccgta	2580
ttggtcagca	accaacaaag	cagcgtaaac	ttggctctggg	gtagcaccac	ctggcatggt	2640
gtggatggtt	tcaccttcag	cacaagcagc	ttcagcaacg	attctcatct	tagctggaac	2700
gtcttccttg	atgtccaatt	gagccaaggt	gattggcaaa	ccaacagcgt	gagacaaagc	2760
agcaacgggt	tcgattttct	caactggagc	gttttccaaa	accaattggg	tcaaggtaac	2820
gaaagcaacc	ttttcaccgt	ggtagtagtg	gtgagcgtct	gggatagcgg	tcaaaccggt	2880
gtgaacagcg	tgagcagcag	ccaaaccacc	agattcgaaa	ccaacaccag	acaagtaggt	2940
gttagcttcg	ataactcttt	ccaaagctgg	ggtaacaacg	tgttgttcag	cagccaacat	3000
agccttttca	ccttcttcca	acaaggtggt	gtaacacaat	tcagccaaaag	ccaaagcagc	3060
ttgggtacac	ttaccaccag	ccatggtggt	agcaccagat	ctagaacaag	ctctagcttc	3120
gaaccaggta	gccaaagcgt	caccgatacc	agctgccaac	aatctagctg	gagcaccagc	3180
aacgatcttg	gtgtcaacga	taaccatggt	tgggttggtt	ggcaacaaca	agtatctgtc	3240
gaattcacct	tcgtcgggtg	agataacaga	caaagcagaa	catggagcgt	cggtagaagc	3300
gatggttgga	gcgatagcaa	ctggaacacc	catgaagtga	gccaaagcct	tagcgggtgtc	3360
caaggtctta	ccaccaccga	tacccaagat	agcaccacat	tgagcggttt	cagcgatacc	3420
tctcaatctg	tcgattttcgt	tttgagaaca	ttcaccaccg	aatggagcga	tttcaacaac	3480
caaaccagcg	tccttgaaaag	acttttcaac	ggtagattga	gcgaaaccca	aaacgaactt	3540
gtcaccaaca	accaaccatc	tttcagccaa	tggcttcaag	tattcaccca	atctgttgat	3600
aacgtcagca	ccttggtatg	acttacctgg	agattggatg	attctgtcca	ttttctgcag	3660
ttaattcagt	aaattttcga	tcttggaag	aaaaaagcag	taagcgtgaa	aaatctaaaa	3720
gotgatgtag	tagaagatcc	tattctttta	caaagattga	ccttttcttt	ttcttcttgg	3780
tttgagttag	aaggggaagg	aagaatacaa	gagagaggaa	aaaaaggaag	ataaaaagag	3840
agcgtgatat	aaatgaatat	atatttaaaca	agagagattg	ggaaggaaaag	gatcaaacaa	3900
acccaaaaat	atttcaaaaa	ggagagagag	aggcaggttt	ggtttcaaaa	cggttttattt	3960
atztatgcaa	gaggacgtgg	aagaaaaaga	agaaggaaga	aaaaaatttg	aaagaaaaaa	4020
acgcgtggcg	ggtaaagaag	aaaatggaaa	atagaggccg	ggtgacagag	aaatattgag	4080
ggttaattgg	aaaatatggt	agggtgaggc	atatgttttt	aagggttttg	aggatccgat	4140
aaggaagaat	gtagggttaa	tgttggtgcat	taattgctgt	ggcagcttac	ccgcttcccc	4200
acacattttac	tagtctcgag	ctcttcaact	caagacgcac	agatattata	acatctgcat	4260
aataggcatt	tgcaagaatt	actcgtgagt	aaggaaagag	tgaggaaacta	tcgcatacct	4320
gcatttaaag	atgccgattt	gggcgcgaat	cctttatttt	ggcttcaccc	tcatactatt	4380
atcagggcca	gaaaaaggaa	gtgtttccct	ccttcttgaa	ttgatgttac	cctcataaag	4440
cacgtggcct	cttatcgaga	aagaaattac	cgtcgtcgt	gatttggttg	caaaaagaac	4500

aaaactgaaa	aaacccagac	acgctcgaact	tcctgtcttc	ctattgattg	cagcttccaa	4560
tttcgtcaca	caacaaggtc	ctagcgacgg	ctcacagggt	ttgtaacaag	caatcgaagg	4620
ttctggaatg	gcgggaaagg	gtttagtacc	acatgctatg	atgccactg	tgatctccag	4680
agcaaagtcc	gttcgatcgt	actgttaact	tctctcttcc	aaacagaatt	gtccgaatcg	4740
tgtgacaaca	acagcctgtt	ctcacacact	cttttcttct	aaccaagggg	gtggtttagt	4800
ttagtagaac	ctcgtgaaac	ttacattttac	atatatataa	acttgcataa	attggtcaat	4860
gcaagaaata	catatttggt	cttttctaat	tcgtagtttt	tcaagttctt	agatgctttc	4920
tttttctctt	ttttacagat	catcaaggaa	gtaattatct	actttttaca	acaaatatat	4980
ctagaaaatg	gctgttacca	acgttgctga	attgaacgct	ttggttgaaa	gggttaagaa	5040
ggctcaaaga	gaatacgcct	ctttcaccca	agaacaagtt	gacaagatct	tcagagctgc	5100
tgttttggt	gctgctgacg	ctagaatccc	attggctaag	atggctgttg	ctgaatctgg	5160
tatgggtatc	gttgaagaca	aggttatcaa	gaaccacttc	gcttctgaat	acatctacaa	5220
cgcttacaag	gacgaaaaga	cctgtggtgt	tttgtoagaa	gacgacacct	tcggtacctat	5280
caccatcgct	gaaccaatcg	gtatcatctg	tggtatcggt	ccaaccacca	acccaacctc	5340
taccgctatc	ttcaagtctt	tgatctcttt	gaagaccaga	aacgctatca	tcttctctcc	5400
acaccaaga	gctaaagacg	ctaccaacaa	ggctgctgac	atcgttttgc	aagctgctat	5460
cgctgctggt	gctccaaagg	acttgatcgg	ttggatcgac	caaccatctg	ttgaattgtc	5520
taacgctttg	atgcaccacc	cagacatcaa	cttgatcttg	gctaccggtg	gtccagggtat	5580
ggttaaggct	gcttactctt	ctggtaagcc	agctatcggt	gttgggtgctg	gtaacacccc	5640
agttgttate	gacgaaaccg	ctgacatcaa	gagagctggt	gcttctgttt	tgatgtctaa	5700
gaccttcgac	aacggtgtta	tctgtgcttc	tgaacaatct	gttgttgttg	ttgactctgt	5760
ttacgacgct	gttagagaaa	gattcgctac	ccacggtggt	tacttggtgc	aaggtaagga	5820
attgaaggct	gttcaagacg	ttatcttgaa	gaacggtgct	ttgaacgctg	ctatcgttgg	5880
tcaaccagct	tacaagatcg	ctgaattagc	tggtttctct	gttccagaaa	acaccaagat	5940
cttgatcggt	gaagttaccg	ttgttgacga	atctgaacca	ttcgctcacg	aaaagttgtc	6000
tccaaccttg	gctatgtaca	gagctaagga	cttcgaagac	gctgttgaaa	aagctgaaaa	6060
gttggttgct	atgggtggta	ttggtcacac	ctcttggttg	tacaccgacc	aagacaacca	6120
accagctaga	gtttcttact	tcggtcaaaa	gatgaagacc	gctagaatct	tgatcaacac	6180
cccagcttct	caagggtggt	tcggtgactt	gtacaacttc	aagttggctc	catctttgac	6240
cttgggttgt	ggttcttggg	gtggtaactc	tatctctgaa	aacgttggtc	caaagcactt	6300
gatcaacaag	aagaccgttg	ctaagagagc	tgaaaacatg	ttgtggcaca	agttgccaaa	6360
atctatctac	ttcagaagag	gttctttgcc	aatcgctttg	gacgaagtta	tcaccgacgg	6420
tcacaagaga	gctttgatcg	ttaccgacag	attcttggtc	aacaacggtt	acgctgacca	6480
aatcacctct	gttttgaaag	ctgctggtgt	tgaaaaccga	gttttcttcg	aagttgaagc	6540
tgacceaaac	ttgtctatcg	ttagaaaggg	tgtgaattg	gctaactctt	tcaagccaga	6600
cgttatcatc	gctttgggtg	gtggttctcc	aatggacgct	gctaagatca	tgtgggttat	6660
gtacgaacac	ccagaaaccc	acttcgaaga	attggctttg	agattcatgg	acatcagaaa	6720
gagaatctac	aagttcccaa	agatgggtgt	taaggotaag	atgatcgctg	ttaccaccac	6780
ctctggtacc	ggttctgaag	ttaccccatt	cgctgttggt	accgacgacg	ctaccggtca	6840
aaagtaccca	ttggctgact	acgctttgac	cccagacatg	gctatcgttg	acgctaactt	6900
ggttatggac	atgccaaagt	ctttgtgtgc	tttcggtggt	ttggacgctg	ttaccacacg	6960
tatggaagct	tacgtttctg	ttttggcttc	tgaattctct	gacggtcaag	ctttgcaagc	7020
tttgaagttg	ttgaagggaat	acttgccagc	ttcttaccac	gaaggttcta	agaaccacgt	7080
tgttagagaa	agagttcact	ctgctgctac	catcgctggt	atcgctttcg	ctaacgcttt	7140
cttgggtggt	tgtcactcta	tggctcacia	gttgggttct	caattccaca	tcccacacgg	7200
tttggctaac	gctttgttga	tctgtaacgt	tatcagatac	aacgctaacg	acaacccaac	7260
caagcaaacc	gctttctctc	aatacgacag	accacaagct	agaagaagat	acgctgaaat	7320
cgctgaccac	ttgggtttgt	ctgctccagg	tgacagaacc	gctgcaaaga	tcgaaaagtt	7380
gttggcttgg	ttggaaacct	tgaaggctga	attgggtatc	ccaaagtcta	tcagagaagc	7440
tgggtgtcaa	gaagctgact	tcttggctaa	cgttgacaag	ttgtctgaag	acgctttcga	7500
cgaccaatgt	accggtgcta	acccaagata	ccattgatc	tctgaattga	agcaaatctt	7560
gttggacacc	tactacggta	gagactacgt	tgaagggtga	accgctgcta	agaaggaagc	7620
tgtccagct	aaggctgaaa	agaaggctaa	gaagtctgct	tagcttaagc	gcgcgaattt	7680
cttatgattt	atgattttta	ttattaaata	agttataaaa	aaaataagtg	tatacaaatt	7740
ttaaagtgac	tcttaggttt	taaaacgaaa	attcttattc	ttgagtaact	ctttcctgta	7800
ggtcagggtg	ctttctcagg	tatagcatga	ggctcgctctt	attgaccaca	cctctaccgg	7860

catgccgagc	aaatgcctgc	aaatcgctcc	ccatttcacc	caattgtaga	tatgctaact	7920
ccagcaatga	gttgatgaat	ctcgggtgtgt	attttatgtc	ctcagaggac	aacacctggt	7980
gtaatcgttc	ttccacacgt	acgaagctta	aaagggcgaa	ttctgcagat	atccatcaca	8040
ctggcgggccg	catgctaget	ccggattatc	gatgataagc	tgtcaaagat	gagaattaat	8100
tccacggact	atagactata	ctagatactc	cgtctactgt	acgatacact	tccgctcagg	8160
tccttgteet	ttaacgaggc	cttaccactc	ttttgttact	ctattgatcc	agctcagcaa	8220
aggcagtggtg	atctaagatt	ctatcttctgc	gatgtagtaa	aactagctag	accgagaaag	8280
agactagaaa	tgcaaaaaggc	acttctacaa	tggctgccat	cattattatc	cgatgtgacg	8340
ctgcagcttc	tcaatgatat	tcgaatacgc	tttgaggaga	tacagcctaa	tatccgacaa	8400
actgtttttac	agattttacga	tcgtacttgt	tacccatcat	tgaattttga	acatccgaac	8460
ctgggagttt	tccctgaaac	agatagtata	tttgaacctg	tataataata	tatagtctag	8520
cgttttacgg	aagacaatgt	atgtatctcg	gttcttgtag	aaactattgc	atctattgca	8580
taggtaatct	tgcacgtcgc	atccccgggt	cattttctgc	gtttccatct	tgcacttcaa	8640
tagcatatct	ttgttaacga	agcatctgtg	cttcattttg	tagaacaaaa	atgcaacgcg	8700
agagcgctaa	tttttcaaac	aaagaatctg	agctgcattt	ttacagaaca	gaaatgcaac	8760
gcgaaagcgc	tattttacca	acgaagaatc	tgtgcttcat	ttttgtaaaa	caaaaatgca	8820
acgcgacgag	agcgctaatt	tttcaaacaa	agaatctgag	ctgcattttt	acagaacaga	8880
aatgcaacgc	gagagcgcta	ttttaccaac	aaagaatcta	tacttctttt	ttgttctaca	8940
aaaatgcac	ccgagagcgc	tatttttcta	acaaagcacc	ttagattact	ttttttctcc	9000
tttgtgcgct	ctataatgca	gtctcttgat	aaactttttg	actgtaggtc	cgttaagggt	9060
agaagaaggc	tactttgggtg	tctattttct	cttccataaa	aaaagcctga	ctccacttcc	9120
cgcgtttact	gattactagc	gaagctgcgg	gtgcattttt	tcaagataaa	ggcatccccg	9180
attatattct	ataccgatgt	ggattgcgca	tactttgtga	acagaaagtg	atagcgttga	9240
tgattcttca	ttgggtcagaa	aattatgaac	ggtttcttct	attttgtctc	tatatactac	9300
gtataggaaa	tgttttacatt	ttcgtattgt	tttcgattca	ctctatgaat	agttcttact	9360
acaatttttt	tgtctaaaga	gtaatactag	agataaacat	aaaaaatgta	gaggtcgagt	9420
ttagatgcaa	gttcaaggag	cgaaagggtg	atgggtaggt	tatataggga	tatagcacag	9480
agatatatat	caaagagata	cttttgagca	atgtttgtgg	aagcgggtatt	cgcaatgccg	9540
gctttccccc	tcaagctcta	aatcgggggc	tccctttagg	gttccgattt	agtgcctttac	9600
ggcacctcga	ccccaaaaaa	cttgattagg	gtgatgggtc	acgtagtggg	ccatcgccct	9660
gatagacggt	ttttcgccct	ttgacgttgg	agtccacggt	ctttaatagt	ggactcttgt	9720
tccaaactgg	aacaacactc	aacctatct	cggctctatc	ttttgattta	taagggtatt	9780
tgccgatttc	ggcctattgg	ttaaaaaatg	agctgattta	acaaaaattt	aacgcgaatt	9840
ttaacaaaat	attaacgttt	acaatttct	gatgcggtat	tttctcctta	cgcctctgtg	9900
cggatatttc	caccgcatat	gatccgtcga	gttcaagaga	aaaaaaaaaga	aaaagcaaaa	9960
agaaaaaagg	aaagcgcgcc	tcgttcagaa	tgacacgtat	agaatgatgc	attaccttgt	10020
catcttcagt	atcatactgt	tcgtatacat	acttactgac	attcataggt	atacatatat	10080
acacatgtat	atatatcgta	tgetgcagct	ttaaataatc	ggtgtcacta	cataagaaca	10140
cctttgggtg	agggaaacatc	gttggtacca	ttggggcgagg	tggtttctct	tatggcaacc	10200
gcaagagcct	tgaacgcact	ctcactacgg	tgatgatcat	tcttgccctg	cagacaatca	10260
acgtggaggg	taattctgct	agcctctgca	aaagctttcaa	gaaaatgcgg	gatcatctcg	10320
caagagagat	ctcctacttt	ctccctttgc	aaaccaagtt	cgacaaactgc	gtacggcctg	10380
ttcgaaagat	ctaccaccgc	tctggaaagt	gcctcatcca	aaggcgcaaa	tctgatcca	10440
aaccttttta	ctccacgcgc	cagtagggcc	tctttaaaag	cttgaccgag	agcaatcccg	10500
cagtcttcag	tggtgtgatg	gtcgtctatg	tgtaagtcac	caatgcactc	aacgattagc	10560
gaccagccgg	aatgcttggc	cagagcatgt	atcatatggt	ccagaaaccc	tatacctgtg	10620
tggacgttaa	tcacttgcca	ttgtgtggcc	tgttctgcta	ctgcttctgc	ctctttttct	10680
gggaagatcg	agtgtctctat	cgctagggga	ccacccttta	aagagatcgc	aatctgaatc	10740
ttggtttcat	ttgtaatacg	ctttactagg	gctttctgct	ctgtcatctt	tgccttcggt	10800
tatcttgctt	gtcatttttt	tagtatatct	ttcgaagaaa	tcacattact	ttatataatg	10860
tataattcat	tatgtgataa	tgccaatcgc	taagaaaaaa	aaagagtcat	ccgctagggtg	10920
gaaaaaaaaa	aatgaaaatc	attaccgagg	cataaaaaaa	tatagagtgt	actagaggag	10980
gccaagagta	atagaaaaag	aaaattgcgg	gaaaggactg	tgttatgact	tccctgacta	11040
atgccgtgtt	caaacgatac	ctggcagtga	ctcctagcgc	tcaccaagct	cttaaaacgg	11100
aattatggtg	cactctcagt	acaatctgct	ctgatgccgc	atagtttaagc	cagccccgac	11160
acccgccaac	acccgctgac	gcgccttgac	gggtctgtct	gctcccgcca	tccgcttaca	11220

gacaagctgt gaccgtctcc gggagctgca tgtgtcagag gttttcaccg tcatcaccga 11280
aacgcgcga 11289

<210> 57
<211> 11102
<212> DNA
<213> Artificial Sequence

<220>
<223> pRN607

<400> 57
agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 60
gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 120
gataacctaca gcgtgagcat tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 180
ggatatccggg aagcggcagg gtccgaacag gagagcgcac gagggagctt ccagggggga 240
acgcctggta tctttatagt cctgtcgggt ttccgccact ctgacttgag cgtcgatttt 300
tgtgatgctc gtcagggggg ccgagcctat ggaaaaacgc cagcaacgcg gcctttttac 360
ggttcctggc cttttgctgg ctttttgcct acatgttctt tctgcgtta tccccctgatt 420
ctgtggataa ccgtattacc gcctttgagt gagctgatac cgctcgccgc agccgaacga 480
ccgagcgcag cgagtcagtg agcaggaag cggagagcg cccaatacgc aaaccgcctc 540
tccccgcgcg ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag 600
cgggcagtga gcgcaacgca attaatgtga gttacctcac tcattaggca ccccaggctt 660
tacactttat gcttcgggt cctatgttgt gtggaattgt gagcggataa caatttcaca 720
caggaaacag ctatgaccat gattacgcca agctcggaat taaccctcac taaagggaac 780
aaaagctggg tacgttttaa acagttgatg agaacccttt tcgcaagttc aaggtgctct 840
aatttttaaa atttttactt ttccgcacac aataaagtct tcacgacgct aaactattag 900
tgcacataat gtagttactt ggacgctgtt caataatgta taaaatttat ttcctttgca 960
ttacgtacat tatataacca aatcttaaaa atatagaaat atgatatgtg tataataata 1020
taagcaaaat ttacgtatct ttgcttataa tatagcttta atgttcttta ggtatatatt 1080
taagagcgat ttgtctcgac ttacaaggcg ctttgaaccc ccttcaaaaa ctccacatagg 1140
cctactgcac caggatcttc tacttgagat gaatcgccga catacgaagc tctgccaaat 1200
ttggcctcga atgtagcagt ggatttagca ccttcctctg cagcttttac cgccttattg 1260
aaatccttag atgcagtaaa ttctttaacg aatggttcta aagcatcaat catggtggat 1320
gatcccttcc ttgcctttgt atatttgtat aaagtatcca atgcaattcc gagtgactta 1380
gccacaattt ccttagtgac gggttcatcc tttgatttac aaacctgaat taatccgtgt 1440
gaaaaacccg acaaaagaat agaataataa ccaccagaag tacctcccat tgagccttca 1500
atgaaatctg aaatttgggc aaccgcctga gataatgagt ccttcgacag cttgtcaagg 1560
ttttcgggtg tgcctttaac tctgccact aaagtgtaac cacaatcacc atcaccaact 1620
tgattgtcta gtcccgtaat gtgcggttcg ctcttgataa cttgttcagc accactcttc 1680
atccactcag caaacttgtc aaagtcatag gtaccgaagg cttttgctgt tacttcatta 1740
tgtaacaagt catcgttaac agatggggca gaagtctttt caaaatctgc aattggccag 1800
cccgggtgct tcgtaaatgc gttcaacaag tctagtactg atttgatctc ctcaaaatca 1860
gattgcaaag cttttgtagc cttagtggcg tttagtaatg tgatactgaa cccattacca 1920
ttgaaggagg tcatcaatgt gccagcaatt gtttgaaccg gggttatgtt gtaattttcc 1980
tttaagaaat ccgtagtttt ggaagtgata gaactaataa caaaattaga aacaccgccg 2040
agattgttaa ctaacaagac aacttcatca tcttcatcaa actttacaaa agctctatcc 2100
ttatcgtttg gatccaatag ttttggtagc atatacttgg agatcaagtc ttcggtagaa 2160
ggaatagggt ctaaaacttt cacaccaggc tcgttatgaa taccataacc caattccatt 2220
tgtttttcgt ttaattcact ttogaatttc ctgccaggaa ctttacaatg gtctaaagaa 2280
gatccaatgg tcaccaaatt gtccgttgata attttagcca ctttagctgt accgtctaag 2340
ccatacttac tagaatatc ttctgcgaag gcacctacaa tcttatgaac caaaacggta 2400
cctgccaatg ctcttctacc aaccatacca cctttttctc tgccaactgc aacatcatca 2460
cctatgacag caacgcggca gttaataccc aaggctcttg ctctctcagc ggacagacca 2520
aaatgcaaaa catcacctgt gtagttcttc acaatcaata aaacgccaga cgcattttca 2580

ttgactaaac	ggattgcatt	taaaatctgt	tttgttgaag	gggatgcaaa	aatttcgcca	2640
accacggcgc	cactcaacat	acccttacca	atgaaaccgg	cgtgtgtagg	ttcatgtcca	2700
ctaccaccac	cagaaattaa	tgcgatcttg	tcggaatcgg	tctttctgaa	gagaattttt	2760
tcttcagggg	ccagcgtaat	ggaggggtta	gcaagggcaa	accctttgag	acttgaattg	2820
actggatctg	tgacttcaaa	cgatttagcg	gacattttct	agatttatgt	atgtgttttt	2880
tgtagttata	gatttaagca	agaaaagaat	acaaacaaaa	aattgaaaaa	gattgattta	2940
gaattaaaaa	gaaaaatatt	tacgtaagaa	gggaaaatag	taaatggtgc	aagttcacta	3000
aactcctaaa	ttatgctgcc	ctttatatcc	cctgttacag	cagccgagcc	aaaggtatat	3060
aggctccttt	gcattagcat	gcgtaacaaa	ccacctgtca	gtttcaaccg	aggtgggtatc	3120
cgagagaatt	gtgtgattgc	tttaattaat	ttcggagaat	ctcacatgcc	actgaagatt	3180
aaaaactgga	tgccagaaaa	gggtgtccca	ggtgtaacat	caatagagga	agctgaaaag	3240
tcttagaacg	ggtaatcttc	caccaacctg	atgggttcct	agatataatc	tcgaagggaa	3300
taagtagggg	gataccgcag	aagtgtctga	atgtattaag	gtcctcacag	tttaaatccc	3360
gctcacacta	acgtaggatt	attataactc	aaaaaaatgg	cattattcta	agtaagttaa	3420
atatccgtaa	tctttaaaca	ctatgtagtt	aggctctcct	cacaatcagt	ccatttgggt	3480
agcacggtec	tcaggacgta	tctattgatg	gattcgtcca	gttccatcac	cattacgctc	3540
ccgttaggaa	cattggtaaa	cgattcaaac	tcttcgtatg	tccatctaaa	ccatttcctc	3600
aggaatactc	tggaataaat	accatgtgta	actaggacaa	caacatctct	gggtcttctc	3660
tcttgccctat	catggaagtg	cctgaataaa	gtctcttgga	aactggcgac	tctgtcatat	3720
acatctgccg	cactttctcc	atgaggggaat	ctgaagaaga	aaccggttta	tcattatcaa	3780
tactgccatt	tcaaagaata	cgtaaataat	taatagtagt	gattttccta	actttattta	3840
gtcaaaaaat	tagcctttta	attctgctgt	aaccgcgtaca	tgcccaaaat	agggggcggg	3900
ttacacagaa	tatataacat	cgtagggtgc	tgggtgaaca	gtttattcct	ggcatccact	3960
aaatataatg	gagcccgctt	tttaagctgg	catccagaaa	aaaaaagaat	cccagcacca	4020
aaatattggt	ttcttcacca	accatcagtt	catagggtcca	ttctcttagc	gcaactacag	4080
agaacagggg	cacaaacagg	caaaaaacgg	gcacaacctc	aatggagtga	tgcaacctgc	4140
ctggagtaaa	tgatgacaca	aggcaattga	cccacgcgat	tatctatctc	attttcttac	4200
accttctatt	accttctgct	ctctctgatt	tggaaaaagc	tgaaaaaaaa	ggttgaaacc	4260
agttccctga	aattattccc	ctacttgact	aataagtata	taaagacggg	aggtattgat	4320
tgtaattctg	taaatctatt	tcttaaaact	cttaaatctc	acttttatag	ttagtctttt	4380
ttttagtttt	aaaacaccaa	gaacttagtt	tcgaataaac	acacataaag	aattcaaaat	4440
gtcgtctgat	agcatcctgt	ctccccta	tacttccgag	ggcttagatt	caagaatcaa	4500
accttcacca	aaaaaggatg	cctctactac	cactaagcca	tcactatgga	aaactactga	4560
gttcaaattc	tactacattg	catttctggt	cgtggttccc	ttgatgttct	atgctggggt	4620
acaagctagt	tcgcccga	atccaaacta	tgcaagatac	gaacgtctcc	tatctcaagg	4680
ttgggtattt	ggcagaaaag	tagacaatag	tgattctcaa	tataggtttt	tcagggacaa	4740
ttttgcgcta	ttgtcagttt	taatgctagt	ccacacttct	ataaaacgca	ttgtacttta	4800
ttcaacaaat	atcactaaat	tgaggtttga	tctgatattt	ggtttgatct	tttttagtggc	4860
cgtctatggt	gtcaattcga	taagaatttt	agcccatatg	ctaattttat	atgccatcgc	4920
ccatgtacta	aagaacttta	gaagaatagc	caccatcagc	atttggattt	atgggtatttc	4980
tacgtttttt	attaacgaca	acttcagagc	atatccattt	ggtaatat	gctctttttt	5040
aagcccattg	gaccattggg	atagagggtat	cattccaaga	tgggatgtct	ttttcaattt	5100
tactcttttg	agagtcttaa	gttacaactt	ggacttctta	gagagggtgg	agaatttaca	5160
aaagaagaaa	agtcctcctc	atgaatcaaa	agaagctaaa	tcagccattt	tgctcaatga	5220
acgtgctaga	ttaactgctg	cacaccccat	acaggactac	agcttaatga	attatattgc	5280
atatgttact	tacacgccac	ttttcattgc	cggcccccatt	ataacattca	atgattatgt	5340
ttaccaatcg	aaacatacct	tgccatcaat	aaatttcaaa	ttcatttttt	actatgcggg	5400
gagattcggt	attgctctct	tatctatgga	gttcatttta	cactttctcc	acgttggtggc	5460
aatctcaaaa	accaaagcgt	gggaaaatga	cacacctttc	cagattttcca	tgattggcctt	5520
atttaatttg	aataattatt	ggctaaaact	actgattccg	tggaggctgt	ttaggctgtg	5580
ggctttgcta	gacggaatcg	atacacctga	aaatatgac	aggtgtgttg	ataacaatta	5640
cagttcacta	gcattctgga	gagcttggca	tagaagctac	aataagtggg	ttgtccgtta	5700
catatatatt	cctctaggtg	gttcaaaaaa	tagagttttg	acatcactag	cagtcttttc	5760
cttcgtagct	atatggcatg	acatcgaaat	aaagttatta	ttatgggggt	ggctaatagt	5820
tttgttctc	ttaccagaaa	tttttgctac	ccaaattttc	tctcattata	ccgacgcagt	5880
ctggtacaga	cacgtttgog	ctgtcgggtg	tgttttcaac	atatgggtta	tgatgatcgc	5940

taatcttttt	ggattctgct	tgggctctga	cggactactaaa	aaattactaa	gcgatatggt	6000
ctgtaccgta	tctgggtttca	aatttgtaat	tttggaagc	gttagtttat	tcacgcagct	6060
acaaataatg	tttgaaatca	gagaagaaga	aaagaggcac	ggaattttacc	taaaatgctg	6120
aggatccct	tttcctttgt	cgatatcatg	taattagtta	tgtcacgctt	acattcacgc	6180
cctcctccca	catccgctct	aaccgaaaag	gaaggagtta	gacaacctga	agtctaggtc	6240
cctattttatt	ttttttaata	gttatgttag	tattaagaac	gttattttata	tttcaaattt	6300
ttcttttttt	tctgtacaaa	cgcgtgtaag	catgtaacat	tatactgaaa	accttgcttg	6360
agaagggtttt	gggacgctcg	aaggcttct	agttctagag	cggcccgccac	cgcgggtggag	6420
ctccaattcg	ccctatagtg	agtcgtatta	caattcactg	gccgtcgttt	tacaacgctg	6480
tgactgggaa	aaccctggcg	ttacccaact	taatcgctt	gcagcacatc	cccccttcgc	6540
cagctggcgt	aatagcgaag	aggcccgcac	cgatcgccct	tccaacagct	tgcgcagcct	6600
gaatggcgaa	tggcgcgacg	cgccctgtag	cggcgcatta	agcgcggcgg	gtgtgggtgt	6660
tacgcgcagc	gtgaccgcta	cacttgccag	cgccttagcg	cccgctcctt	tcgctttctt	6720
cccttccttt	ctcgccacgt	tcgcggtag	tgtagaacct	gaacaagggt	tactaaaaat	6780
ccgtaaagaa	cttcaattgt	acgccaactt	aaggcctcgt	ccccgcggg	tcaccggcc	6840
agcgacatgg	aggcccagaa	tacctcctt	gacagtcttg	acgtgcgcag	ctcaggggca	6900
tgatgtgact	gtcgcccgtg	catttagccc	atacatcccc	atgtataatc	atttgcctcc	6960
atacattttg	atggccgcac	ggcgcgaagc	aaaaattacg	gtcctcctct	gcagacctgc	7020
gagcagggaa	acgtccctt	cacagacgcg	tgaattgtcc	ccacgcgcgc	cccctgtaga	7080
gaaatataaa	aggttaggat	ttgccactga	ggttcttctt	tcataactt	ccttttataa	7140
tcttgctagg	atacagttct	cacatcacat	ccgaacataa	acaacctatg	gtaaaaagcc	7200
tgaactcacc	gcgacgtctg	tcgagaagtt	tctgatcgaa	aagttcgaca	gcgtctccga	7260
cctgatgcag	ctctcggagg	gcgaagaatc	tcgtgctttc	agcttcgatg	taggagggcg	7320
tggatatgtc	ctgcgggtaa	atagctgcgc	cgatgggttt	tacaaagatc	gttatgttta	7380
tcggcacttt	gcateggccg	cgtcccgat	tcgggaagtg	cttgacattg	gggaattcag	7440
cgagagcctg	acctattgca	tctcccgcgc	tgcacagggt	gtcacgttgc	aagacctgcc	7500
tgaacccgaa	ctgcccgcgt	ttctgcagcc	ggtcgcggag	gccatggatg	cgatcgctgc	7560
ggccgatctt	agccagacga	gcgggttcgg	cccatccgga	ccgcaaggaa	tcgggtcaata	7620
cactacatgg	cgtgatttca	tatgcgcgat	tgtgatccc	catgtgtatc	actggcaaac	7680
tgtgatggac	gacaccgtca	gtgcgtccgt	cgcgcaggct	ctcgatgagc	tgatgctttg	7740
ggccgaggac	tgccccgaag	tccggcacct	cgtgcacgcg	gatttcggct	ccaacaatgt	7800
cctgacggac	aatggccgca	taacagcggg	cattgactgg	agcgaggcga	tgttcgggga	7860
ttcccaatac	gaggtcgcga	acatcttctt	ctggaggccg	tggttggctt	gtatggagca	7920
gcagacgcgc	tacttcgagc	ggaggcatcc	ggagcttgca	ggatcgccgc	ggctccgggc	7980
gtatatgctc	cgcattggtc	ttgaccaact	ctatcagagc	ttggttgacg	gcaatttcga	8040
tgatgcagct	tgggcgcagg	gtcgatgcga	cgcaatcgtc	cgatccggag	ccgggactgt	8100
cgggcgtaca	caaatcgccc	gcagaagcgc	ggccgtctgg	accgatggct	gtgtagaagt	8160
actcgccgat	agtggaaacc	gacgccccag	cactcgtccg	agggcaaagg	aataatcagt	8220
actgacaata	aaaagattct	tgttttcaag	aaactgtcat	ttgtatagtt	tttttatatt	8280
gtagttgttc	tattttaatc	aatgtttagc	gtgattttata	ttttttttcg	cctcgacatc	8340
atctgcccag	atgcgaagtt	aagtgcgcag	aaagtaatat	catgcgtcaa	tcgtatgtga	8400
atgctggctg	ctatactgct	gtcgattcga	tactaaagcc	gccatccagt	gtcgacggat	8460
cctaggtgta	cataaacttt	ataaatgaaa	ttcataatag	aaacgacacg	aaattacaaa	8520
atggaatatg	ttcatagggg	agacgaaact	atatacgcaa	tctacataca	tttatcaaga	8580
aggagaaaaa	ggaggatagt	aaaggaatac	aggtaagcaa	attgatacta	atggctcaac	8640
gtgataagga	aaaagaattg	cactttaaca	ttaatatgta	caaggaggag	ggcaccacac	8700
aaaaagttag	gtgtaacaga	aatcatgaa	actacgatc	ctaatttgat	attggaggat	8760
tttctctaaa	aaaaaaaaaa	tacaacaaat	aaaaaacact	caatgacctg	accatttgat	8820
ggagttttaag	tcaatacctt	cttgaaccat	ttcccataat	ggtgaaagtt	ccctcaagaa	8880
ttttactctg	tcagaaacgg	ccttacgacg	tagtcgatat	ggtgcactct	cagtacaatc	8940
tgtctctgatg	ccgcatagtt	aagccagccc	cgacaccgcg	caacaccgcg	tgacgcgccc	9000
tgacgggctt	gtctgctccc	ggcatccgct	tacagacaag	ctgtgaccgt	ctccgggagc	9060
tgcattgtgc	agaggttttc	accgtcatca	ccgaaacgcg	cgagacgaaa	gggcctcgtg	9120
atacgccctat	ttttataggt	taatgtcatg	ataataatgg	tttcttagga	cggatcgctt	9180
gcctgtaact	tacacgcgcc	tcgtatcttt	taatgatgga	ataatttggg	aatttactct	9240
gtgtttattt	atttttatgt	tttgtatgtg	gatttttagaa	agtaaataaa	gaaggtagaa	9300

gagttacgga	atgaagaaaa	aaaaataaac	aaagggtttaa	aaaattttcaa	caaaaagcgt	9360
actttacata	tatatattatt	agacaagaaa	agcagattaa	atagatatatac	attcagattaa	9420
cgataagtaa	aatgtaaaaat	cacaggattt	togtgtgtgg	tcttctacac	agacaagatg	9480
aaacaattcg	gcattaatac	ctgagagcag	gaagagcaag	ataaaaaggta	gtatttggtg	9540
gcgatecccc	tagagtcttt	tacatcttcg	gaaaacaaaa	actatTTTTT	ctttaatttc	9600
TTTTTTTact	ttctatTTTT	aatttatata	tttatattaa	aaaattttaa	ttataattat	9660
ttttatagca	cgtgatgaaa	aggacccagg	tggcactttt	cggggaaatg	tgcgcggaac	9720
ccctatTTgt	ttatTTTTct	aaatacatte	aaatatgtat	ccgctcatga	gacaataacc	9780
ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	atttccgtgt	9840
cgcctttatt	ccctTTTTtg	cggcattttg	ccttcctggt	tttgctcacc	cagaaacgct	9900
ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	gtgggttaca	tcgaactgga	9960
tctcaacagc	ggtaagatec	ttgagagttt	tgcgccgaa	gaacgttttc	caatgatgag	10020
cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	attgacgccg	ggcaagagca	10080
actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	gagtactcac	cagtcacaga	10140
aaagcatctt	acggatggca	tgacagtaag	agaattatgc	agtgctgcca	taaccatgag	10200
tgataacact	gcggccaact	tacttctgac	aacgatcgga	ggaccgaagg	agctaaccgc	10260
TTTTTTTcac	aacatggggg	atcatgtaac	tgccttgat	cgttggaac	cggagctgaa	10320
tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	gtagcaatgg	caacaacggt	10380
gcgcaaacta	ttaactggcg	aactacttac	tctagcttcc	cggcaacaat	taatagactg	10440
gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	gcccttccgg	ctggctgggt	10500
tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	ggtatcattg	cagcactggg	10560
gccagatggt	aagccctccc	gtatcgtagt	tatctacacg	acgggcagtc	aggcaactat	10620
ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	ctgattaagc	attggtaact	10680
gtcagaccaa	gtttactcat	atatacttta	gattgattta	aaacttcatt	tttaatttaa	10740
aaggatctag	gtgaagatec	TTTTtgataa	tctcatgacc	aaaatccctt	aacgtgagtt	10800
ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	ggatcttctt	gagatccttt	10860
TTTTctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	ccgctaccag	cggtggtttg	10920
tttgccggat	caagagctac	caactctttt	tccgaaggta	actggettca	gcagagcgca	10980
gataccaaat	actgtccttc	tagtgtagcc	gtagttaggc	caccacttca	agaactctgt	11040
agcacgcct	acatacctcg	ctctgctaata	cctgttacca	gtggctgctg	ccagtggcga	11100
ta						11102

CLAIMS:

1. A yeast cell comprising:

a) an exogenous gene coding for an enzyme with the ability to convert pyruvate and coenzyme-A into formate and acetyl-CoA;

5 b) a genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in the cell;

c) an exogenous gene coding for an enzyme with acetaldehyde dehydrogenase activity, which gene confers to the cell the ability to reduce acetylCoA into acetaldehyde; and,

10 d) a genetic modification that increases the specific activity of NAD^+ -linked glycerol dehydrogenase.

2. A yeast cell according to claim 1, wherein the cell further comprises a genetic modification selected from the group consisting of:

a) a genetic modification that increases the specific activity of dihydroxyacetone kinase;

15 b) a genetic modification that increases transport of glycerol into the cell; an exogenous gene coding for an enzyme for activation of the pyruvate formate lyase;

c) a genetic modification that increases the specific acetyl-CoA synthetase activity in the cell; and,

20 d) a genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell, while the yeast cell comprises a functional high-osmolarity glycerol response pathway.

3. A yeast cell according to claim 1 or 2, wherein the cell comprises at least one of:

i) an exogenous xylose isomerase gene, which gene confers to the cell the ability to isomerise xylose into xylulose; and,

ii) exogenous genes coding for a L-arabinose isomerase, a L-ribulokinase and a L-ribulose-5-phosphate 4-epimerase, which genes together confer to the cell the ability to convert L-arabinose into D-xylulose 5-phosphate.

4. A yeast cell according to any one of claims 1 - 3, wherein at least one of:

5 a) the genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in the cell comprises the inactivation of at least one endogenous copy of a gene encoding a formate dehydrogenase in the cell's genome;

b) the genetic modification that increases the specific activity of the NAD^+ -linked glycerol dehydrogenase is expression of a heterologous gene encoding an NAD^+ -linked glycerol dehydrogenase;

c) the genetic modification that increases transport of glycerol into the cell is overexpression of a nucleotide sequence encoding at least one of a glycerol uptake protein and a glycerol channel;

d) the genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell comprises the inactivation of at least one endogenous copy of a gene encoding a glycerol 3-phosphate dehydrogenase in the cell's genome

e) the genetic modification that increases the specific acetyl-CoA synthetase activity is overexpression of a nucleotide sequence encoding an acetyl-CoA synthetase; and,

20 f) the genetic modification that increases the specific activity of dihydroxyacetone kinase is overexpression of a nucleotide sequence encoding a dihydroxyacetone kinase.

5. A yeast cell according to any one of claims 1 - 4, wherein:

a) the exogenous gene coding for an enzyme with pyruvate formate lyase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 57% amino acid sequence identity with SEQ ID NO: 1;

b) the exogenous gene coding for the pyruvate formate lyase activating enzyme comprises a nucleotide sequence coding for an amino acid sequence with at least 70% amino acid sequence identity with SEQ ID NO: 3;

5 c) the gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is a gene encoding a formate dehydrogenase having an amino acid sequence with at least 70% sequence identity to at least one of SEQ ID NO: 5 and 6;

d) the exogenous gene coding for the enzyme with acetaldehyde dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least one
10 of:

i) at least 64% amino acid sequence identity with SEQ ID NO: 7,

ii) at least 76% amino acid sequence identity with SEQ ID NO: 9 and

iii) at least 61% amino acid sequence identity with SEQ ID NO: 11;

e) the nucleotide sequence encoding the NAD⁺-linked glycerol dehydrogenase
15 comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with SEQ ID NO: 49,

f) the nucleotide sequence encoding the dihydroxyacetone kinase comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with at least one of SEQ ID NO: 14, 15 and 52;

20 g) the nucleotide sequence encoding the glycerol uptake protein comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with at least one of SEQ ID NO: 16 and 17, and wherein the nucleotide sequence encoding the glycerol channel comprises a nucleotide sequence coding for an amino acid sequence with at least 30% amino acid sequence identity with the amino acid sequence
25 between amino acids 250 and 530 of SEQ ID NO: 18;

h) the nucleotide sequence encoding the acetyl-CoA synthetase comprises an amino acid sequence with at least 70% amino acid sequence identity with at least one of SEQ ID NO's: 19 and 20; and,

i) the gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is a gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 21, while the cell has at least one functional copy of a endogenous gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 22.

6. A yeast cell according to any one of claims 1-5, wherein the yeast cell comprises at least one further genetic modification that results in a characteristic selected from the group consisting of:

- a) increased xylulose kinase specific activity;
- b) increased flux of the pentose phosphate pathway
- c) reduced unspecific aldose reductase specific activity
- d) increased transport of at least one of xylose and arabinose into the host cell;
- e) decreased sensitivity to catabolite repression;
- f) increased tolerance to ethanol, osmolarity or organic acids; and,
- g) reduced production of by-products.

7. A yeast cell according to any one of claims 1-6, wherein the yeast cell is of a genus selected from the group consisting of *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*.

8. A yeast cell according to claim 7, wherein the yeast cell belongs to a species selected from the group consisting of *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.

9. A yeast cell according to any one of claims 1-8 wherein at least one of the nucleotide sequences coding for the enzyme with pyruvate formate lyase activity, the pyruvate formate lyase activating enzyme, the enzyme with acetaldehyde dehydrogenase activity, the xylose isomerase, the L-arabinose isomerase, the L-ribulokinase and the L-ribulose-5-phosphate 4-epimerase is a nucleotide sequence the codon usage of which has been optimised for expression in the yeast cell.
10. A process for producing a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, a butanol and an isoprenoid-derived product, whereby the process comprises fermenting a medium with a yeast cell as defined in any one of claims 1 - 9, whereby the medium contains or is fed with a source of glycerol and with a source of at least one of a hexose and a pentose, whereby the yeast cell ferments the glycerol and the at least one of the hexose and pentose to the fermentation product and formate.
11. A process according to claim 10, wherein the fermentation product is ethanol.
12. A process according to claim 10 or 11, wherein the medium contains or is fed with a lignocellulosic hydrolysate.
13. A process according to any one of the claims 10-12, wherein the yeast cell ferments under anaerobic conditions.
14. A process according to any one of the claims 10-12, wherein the pH of the medium is regulated during the fermentation process to maintain a concentration of undissociated formic acid that is not higher than 20.0 mM.
15. A process according to any one of claims 10-14, further comprising recovering at least one of the fermentation product and formate.