

TITLERECOMBINANT BACTERIA HAVING THE ABILITY
TO METABOLIZE SUCROSE

5

FIELD OF THE INVENTION

The invention relates to the fields of microbiology and molecular biology. More specifically, recombinant bacteria having the ability to use sucrose as a carbon source and methods of utilizing such recombinant bacteria to produce products such as glycerol and glycerol-derived products are provided

10

BACKGROUND OF THE INVENTION

Many commercially useful microorganisms use glucose as their main carbohydrate source. However, a disadvantage of the use of glucose by microorganisms developed for production of commercially desirable products is the high cost of glucose. The use of sucrose and mixed feedstocks containing sucrose and other sugars as carbohydrate sources for microbial production systems would be more commercially desirable because these materials are usually readily available at a lower cost.

20

A production microorganism can function more efficiently when it can utilize any sucrose present in a mixed feedstock. Therefore, when a production microorganism does not have the ability to utilize sucrose efficiently as a major carbon source, it cannot operate as efficiently. For example, bacterial cells typically show preferential sugar use, with glucose being the most preferred. In artificial media containing mixtures of sugars, glucose is typically metabolized to its entirety ahead of other sugars. Moreover, many bacteria lack the ability to utilize sucrose. For example, less than 50% of *Escherichia coli* (*E. coli*) strains have the ability to utilize sucrose. Thus, when a production microorganism cannot utilize sucrose as a carbohydrate source, it is desirable to engineer the microorganism so that it can utilize sucrose.

25

30

Recombinant bacteria that have been engineered to utilize sucrose by incorporation of sucrose utilization genes have been reported. For example, Livshits et al. (U.S. Patent No. 6,960,455) describe the production of amino acids using *Escherichia coli* strains containing genes
5 encoding a metabolic pathway for sucrose utilization. Additionally, Olson et al. (*Appl. Microbiol. Biotechnol.* 74:1031-1040, 2007) describe *Escherichia coli* strains carrying genes responsible for sucrose degradation, which produce L-tyrosine or L-phenylalanine using sucrose as a carbon source. However, there is a need for bacterial strains that
10 are engineered to utilize sucrose using new sucrose utilization genes and that have an improved ability to utilize sucrose. Additionally, there is a need for bacterial strains that are capable of producing glycerol and glycerol-derived products using sucrose as carbon source.

15 SUMMARY OF THE INVENTION

In one aspect, there is provided a recombinant bacterium comprising in its genome or on at least one recombinant construct:

- (a) a nucleotide sequence encoding a polypeptide having sucrose transporter activity, the polypeptide having at least 95%
20 sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:24; and
- (b) a nucleotide sequence encoding a polypeptide having sucrose hydrolase activity, the polypeptide having at least 95%
25 sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:26;

wherein (a) and (b) are each operably linked to the same or a different promoter, further wherein said recombinant bacterium is capable of
30 metabolizing sucrose.

In one embodiment, the recombinant bacterium produces 1,3-propanediol, glycerol, and/or 3-hydroxypropionic acid.

In another aspect, there is provided a process for making glycerol, 1,3-propanediol and/or 3-hydroxypropionic acid from sucrose comprising:

- a) culturing the recombinant bacterium that produces 1,3-propanediol, glycerol, and/or 3-hydroxypropionic acid, disclosed herein, in the presence of sucrose; and
- b) optionally, recovering the glycerol, 1,3-propanediol and/or 3-hydroxypropionic acid produced.

BRIEF SEQUENCE DESCRIPTIONS

The following sequences conform with 37 C.F.R. 1.821 1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (2009) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

Table A
Summary of Gene and Protein SEQ ID Numbers

Gene	Coding Sequence SEQ ID NO:	Encoded Protein SEQ ID NO:
<i>GPD1</i> from <i>Saccharomyces cerevisiae</i>	1	2
<i>GPD2</i> from <i>Saccharomyces cerevisiae</i>	3	4
<i>GPP1</i> from <i>Saccharomyces cerevisiae</i>	5	6
<i>GPP2</i> from <i>Saccharomyces cerevisiae</i>	7	8
<i>dhaB1</i> from <i>Klebsiella pneumoniae</i>	9	10
<i>dhaB2</i> from <i>Klebsiella pneumoniae</i>	11	12
<i>dhaB3</i> from <i>Klebsiella pneumoniae</i>	13	14
<i>aldB</i> from <i>Escherichia coli</i>	15	16
<i>aldA</i> from <i>Escherichia coli</i>	17	18
<i>aldH</i> from <i>Escherichia coli</i>	19	20
<i>galP</i> from <i>Escherichia coli</i>	21	22
<i>lacY</i> from <i>Pseudomonas fluorescens</i> Pf5	23	24
<i>scrB</i> from <i>Pseudomonas fluorescens</i> Pf5	25	26
<i>cscB</i> from <i>Escherichia coli</i> EC3132	27	28
<i>cscA</i> from <i>Escherichia coli</i> EC3132	29	30
<i>scrK</i> from <i>Agrobacterium tumefaciens</i>	31	32
<i>scrK</i> from <i>Streptococcus mutans</i>	33	34
<i>scrK</i> From <i>Escherichia coli</i>	35	36
<i>scrK</i> from <i>Klebsiella pneumoniae</i>	37	38
<i>cscK</i> from <i>Escherichia coli</i>	39	40
<i>cscK</i> from <i>Enterococcus faecalis</i>	41	42
<i>HXK1</i> from <i>Saccharomyces cerevisiae</i>	43	44
<i>HXK2</i> from <i>Saccharomyces cerevisiae</i>	45	46
<i>dhaT</i> from <i>Klebsiella pneumoniae</i>	47	48
<i>dhaX</i> from <i>Klebsiella pneumoniae</i>	49	50

SEQ ID NO:51 is the nucleotide sequence of plasmid pSYCO101.

5 SEQ ID NO:52 is the nucleotide sequence of plasmid pSYCO103.

SEQ ID NO:53 is the nucleotide sequence of plasmid pSYCO106.

SEQ ID NO:54 is the nucleotide sequence of plasmid pSYCO109.

SEQ ID NO:55 is the nucleotide sequence of plasmid
pSYCO400/AGRO.

10 SEQ ID NOs:56-61 are the nucleotide sequences of primers used
in the Examples herein.

DETAILED DESCRIPTION

The disclosure of each reference set forth herein is hereby

incorporated by reference in its entirety.

As used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes one or more
5 cells and equivalents thereof known to those skilled in the art, and so forth.

In the context of this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Open reading frame" is abbreviated as "ORF".

"Polymerase chain reaction" is abbreviated as "PCR".

10 "American Type Culture Collection" is abbreviated as "ATCC".

The term "recombinant glycerol-producing bacterium" refers to a bacterium that has been genetically engineered to be capable of producing glycerol and/or glycerol-derived products.

The term "polypeptide having sucrose transporter activity" refers to
15 a polypeptide that is capable of mediating the transport of sucrose into microbial cells.

The term "polypeptide having fructokinase activity" refers to a polypeptide that has the ability to catalyze the conversion of D-fructose + ATP to fructose-phosphate + ADP. Typical of fructokinase is EC
20 2.7.1.4. Enzymes that have some ability to phosphorylate fructose, whether or not this activity is their predominant activity, may be referred to as a fructokinase. Abbreviations used for genes encoding fructokinases and proteins having fructokinase activity include, for example, "*Frk*", "*scrK*", "*cscK*", "*FK*", and "*KHK*". Fructokinase is
25 encoded by the *scrK* gene in *Agrobacterium tumefaciens* and *Streptococcus mutans*; and by the *cscK* gene in certain *Escherichia coli* strains.

The term "polypeptide having sucrose hydrolase activity" refers to a polypeptide that has the ability to catalyze the hydrolysis of sucrose to
30 produce glucose and fructose. Such polypeptides are often referred to as "invertases" or " β -fructofuranosidases".

The terms "glycerol derivative" and "glycerol-derived products"

are used interchangeably herein and refer to a compound that is synthesized from glycerol or in a pathway that includes glycerol. Examples of such products include 3-hydroxypropionic acid, methylglyoxal, 1,2-propanediol, and 1,3-propanediol.

5 The term "microbial product" refers to a product that is microbially produced, i.e., the result of a microorganism metabolizing a substance. The product may be naturally produced by the microorganism, or the microorganism may be genetically engineered to produce the product.

10 The terms "phosphoenolpyruvate-sugar phosphotransferase system", "PTS system", and "PTS" are used interchangeably herein and refer to the phosphoenolpyruvate-dependent sugar uptake system.

15 The terms "phosphocarrier protein HPr" and "PtsH" refer to the phosphocarrier protein encoded by *ptsH* in *E. coli*. The terms "phosphoenolpyruvate-protein phosphotransferase" and "PtsI" refer to the phosphotransferase, EC 2.7.3.9, encoded by *ptsI* in *E. coli*. The terms "glucose-specific IIA component", and "Crr" refer to enzymes designated as EC 2.7.1.69, encoded by *crr* in *E. coli*. PtsH, PtsI, and Crr comprise the PTS system.

20 The term "PTS minus" refers to a microorganism that does not contain a PTS system in its native state or a microorganism in which the PTS system has been inactivated through the inactivation of a PTS gene.

25 The terms "glycerol-3-phosphate dehydrogenase" and "G3PDH" refer to a polypeptide responsible for an enzyme activity that catalyzes the conversion of dihydroxyacetone phosphate (DHAP) to glycerol 3-phosphate (G3P). *In vivo* G3PDH may be NAD- or NADP-dependent. When specifically referring to a cofactor specific glycerol-3-phosphate dehydrogenase, the terms "NAD-dependent glycerol-3-phosphate dehydrogenase" and "NADP-dependent glycerol-3-phosphate dehydrogenase" will be used. As it is generally the case that NAD-
30 dependent and NADP-dependent glycerol-3-phosphate dehydrogenases are able to use NAD and NADP interchangeably (for example by the enzyme encoded by *gpsA*), the terms NAD-dependent and NADP-

dependent glycerol-3-phosphate dehydrogenase will be used interchangeably. The NAD-dependent enzyme (EC 1.1.1.8) is encoded, for example, by several genes including *GPD1*, also referred to herein as *DAR1* (coding sequence set forth in SEQ ID NO:1; encoded protein
5 sequence set forth in SEQ ID NO:2), or *GPD2* (coding sequence set forth in SEQ ID NO:3; encoded protein sequence set forth in SEQ ID NO:4), or *GPD3*. The NADP-dependent enzyme (EC 1.1.1.94) is encoded, for example, by *gpsA*.

The terms "glycerol 3-phosphatase", "sn-glycerol 3-phosphatase",
10 "D,L-glycerol phosphatase", and "G3P phosphatase" refer to a polypeptide having an enzymatic activity that is capable of catalyzing the conversion of glycerol 3-phosphate and water to glycerol and inorganic phosphate. G3P phosphatase is encoded, for example, by *GPP1* (coding sequence set forth in SEQ ID NO:5; encoded protein sequence set forth in SEQ ID
15 NO:6), or *GPP2* (coding sequence set forth in SEQ ID NO:7; encoded protein sequence set forth in SEQ ID NO:8).

The term "glycerol dehydratase" or "dehydratase enzyme" refers to a polypeptide having enzyme activity that is capable of catalyzing the conversion of a glycerol molecule to the product,
20 3-hydroxypropionaldehyde (3-HPA).

For the purposes of the present invention the dehydratase enzymes include a glycerol dehydratase (E.C. 4.2.1.30) and a diol dehydratase (E.C. 4.2.1.28) having preferred substrates of glycerol and 1,2-propanediol, respectively. Genes for dehydratase enzymes have been
25 identified in *Klebsiella pneumoniae*, *Citrobacter freundii*, *Clostridium pasteurianum*, *Salmonella typhimurium*, *Klebsiella oxytoca*, and *Lactobacillus reuteri*, among others. In each case, the dehydratase is composed of three subunits: the large or "α" subunit, the medium or "β" subunit, and the small or "γ" subunit. The genes are also described in, for
30 example, Daniel *et al.* (*FEMS Microbiol. Rev.* 22, 553 (1999)) and Toraya and Mori (*J. Biol. Chem.* 274, 3372 (1999)). Genes encoding the large or "α" (alpha) subunit of glycerol dehydratase include *dhaB1* (coding

sequence set forth in SEQ ID NO:9, encoded protein sequence set forth in SEQ ID NO:10), *gldA* and *dhaB*; genes encoding the medium or “ β ” (beta) subunit include *dhaB2* (coding sequence set forth in SEQ ID NO:11, encoded protein sequence set forth in SEQ ID NO:12), *gldB*, and *dhaC*;
5 genes encoding the small or “ γ ” (gamma) subunit include *dhaB3* (coding sequence set forth in SEQ ID NO:13, encoded protein sequence set forth in SEQ ID NO:14), *gldC*, and *dhaE*. Other genes encoding the large or “ α ” subunit of diol dehydratase include *pduC* and *pddA*; other genes encoding the medium or “ β ” subunit include *pduD* and *pddB*; and other
10 genes encoding the small or “ γ ” subunit include *pduE* and *pddC*.

Glycerol and diol dehydratases are subject to mechanism-based suicide inactivation by glycerol and some other substrates (Daniel *et al.*, *FEMS Microbiol. Rev.* 22, 553 (1999)). The term “dehydratase reactivation factor” refers to those proteins responsible for reactivating
15 the dehydratase activity. The terms “dehydratase reactivating activity”, “reactivating the dehydratase activity” and “regenerating the dehydratase activity” are used interchangeably and refer to the phenomenon of converting a dehydratase not capable of catalysis of a reaction to one capable of catalysis of a reaction or to the phenomenon
20 of inhibiting the inactivation of a dehydratase or the phenomenon of extending the useful half-life of the dehydratase enzyme *in vivo*. Two proteins have been identified as being involved as the dehydratase reactivation factor (see, e.g., U.S. Patent No. 6,013,494 and references therein; Daniel *et al.*, *supra*; Toraya and Mori, *J. Biol. Chem.* 274, 3372 (1999); and Tobimatsu *et al.*, *J. Bacteriol.* 181, 4110 (1999)). Genes encoding one of the proteins include, for example, *orfZ*, *dhaB4*, *gdrA*, *pduG* and *ddrA*. Genes encoding the second of the two proteins include, for example, *orfX*, *orf2b*, *gdrB*, *pduH* and *ddrB*.

The terms “1,3-propanediol oxidoreductase”, “1,3-propanediol dehydrogenase” and “DhaT” are used interchangeably herein and refer
30 to the polypeptide(s) having an enzymatic activity that is capable of

catalyzing the interconversion of 3-HPA and 1,3-propanediol provided the gene(s) encoding such activity is found to be physically or transcriptionally linked to a dehydratase enzyme in its natural (i.e., wild type) setting; for example, the gene is found within a *dha* regulon as is the case with *dhaT* from *Klebsiella pneumoniae*. Genes encoding a 1,3-propanediol oxidoreductase include, but are not limited to, *dhaT* from *Klebsiella pneumoniae*, *Citrobacter freundii*, and *Clostridium pasteurianum*. Each of these genes encode a polypeptide belonging to the family of type III alcohol dehydrogenases, which exhibits a conserved iron-binding motif, and has a preference for the NAD⁺/NADH linked interconversion of 3-HPA and 1,3-propanediol (Johnson and Lin, *J. Bacteriol.* 169, 2050 (1987); Daniel *et al.*, *J. Bacteriol.* 177, 2151 (1995); and Leurs *et al.*, *FEMS Microbiol. Lett.* 154, 337 (1997)). Enzymes with similar physical properties have been isolated from *Lactobacillus brevis* and *Lactobacillus buchneri* (Veiga da Dunha and Foster, *Appl. Environ. Microbiol.* 58, 2005 (1992)).

The term "*dha* regulon" refers to a set of associated polynucleotides or open reading frames encoding polypeptides having various biological activities, including but not limited to a dehydratase activity, a reactivation activity, and a 1,3-propanediol oxidoreductase. Typically a *dha* regulon comprises the open reading frames *dhaR*, *orfY*, *dhaT*, *orfX*, *orfW*, *dhaB1*, *dhaB2*, *dhaB3*, and *orfZ* as described in U.S. Patent No. 7,371,558.

The terms "aldehyde dehydrogenase" and "Ald" refer to a polypeptide that catalyzes the conversion of an aldehyde to a carboxylic acid. Aldehyde dehydrogenases may use a redox cofactor such as NAD, NADP, FAD, or PQQ. Typical of aldehyde dehydrogenases is EC 1.2.1.3 (NAD-dependent); EC 1.2.1.4 (NADP-dependent); EC 1.2.99.3 (PQQ-dependent); or EC 1.2.99.7 (FAD-dependent). An example of an NADP-dependent aldehyde dehydrogenase is AldB (SEQ ID NO:16), encoded by the *E. coli* gene *aldB* (coding sequence set forth in SEQ ID NO:15). Examples of NAD-dependent aldehyde dehydrogenases include AldA

(SEQ ID NO:18), encoded by the *E. coli* gene *aldA* (coding sequence set forth in SEQ ID NO:17); and AldH (SEQ ID NO:20), encoded by the *E. coli* gene *aldH* (coding sequence set forth in SEQ ID NO:19).

The terms "glucokinase" and "Glc" are used interchangeably
5 herein and refer to a protein that catalyzes the conversion of D-glucose + ATP to glucose 6-phosphate + ADP. Typical of glucokinase is EC 2.7.1.2. Glucokinase is encoded by *glk* in *E. coli*.

The terms "phosphoenolpyruvate carboxylase" and "Ppc" are used interchangeably herein and refer to a protein that catalyzes the
10 conversion of phosphoenolpyruvate + H₂O + CO₂ to phosphate + oxaloacetic acid. Typical of phosphoenolpyruvate carboxylase is EC 4.1.1.31. Phosphoenolpyruvate carboxylase is encoded by *ppc* in *E. coli*.

The terms "glyceraldehyde-3-phosphate dehydrogenase" and
15 "GapA" are used interchangeably herein and refer to a protein having an enzymatic activity capable of catalyzing the conversion of glyceraldehyde 3-phosphate + phosphate + NAD⁺ to 3-phospho-D-glyceroyl-phosphate + NADH + H⁺. Typical of glyceraldehyde-3-phosphate dehydrogenase is EC 1.2.1.12. Glyceraldehyde-3-phosphate
20 dehydrogenase is encoded by *gapA* in *E. coli*.

The terms "aerobic respiration control protein" and "ArcA" are used interchangeably herein and refer to a global regulatory protein. The aerobic respiration control protein is encoded by *arcA* in *E. coli*.

The terms "methylglyoxal synthase" and "MgsA" are used
25 interchangeably herein and refer to a protein having an enzymatic activity capable of catalyzing the conversion of dihydroxyacetone phosphate to methylglyoxal + phosphate. Typical of methylglyoxal synthase is EC 4.2.3.3. Methylglyoxal synthase is encoded by *mgsA* in *E. coli*.

The terms "phosphogluconate dehydratase" and "Edd" are used
30 interchangeably herein and refer to a protein having an enzymatic activity capable of catalyzing the conversion of 6-phospho-gluconate to 2-keto-3-deoxy-6-phospho-gluconate + H₂O. Typical of

phosphogluconate dehydratase is EC 4.2.1.12. Phosphogluconate dehydratase is encoded by *edd* in *E. coli*.

The term "YciK" refers to a putative enzyme encoded by *yciK* which is translationally coupled to *btuR*, the gene encoding Cob(I)alamin adenosyltransferase in *E. coli*.

The term "cob(I)alamin adenosyltransferase" refers to an enzyme capable of transferring a deoxyadenosyl moiety from ATP to the reduced corrinoid. Typical of cob(I)alamin adenosyltransferase is EC 2.5.1.17. Cob(I)alamin adenosyltransferase is encoded by the gene "*btuR*" in *E. coli*, "*cobA*" in *Salmonella typhimurium*, and "*cobO*" in *Pseudomonas denitrificans*.

The terms "galactose-proton symporter" and "GalP" are used interchangeably herein and refer to a protein having an enzymatic activity capable of transporting a sugar and a proton from the periplasm to the cytoplasm. D-glucose is a preferred substrate for GalP. Galactose-proton symporter is encoded by *galP* in *Escherichia coli* (coding sequence set forth in SEQ ID NO:21, encoded protein sequence set forth in SEQ ID NO:22).

The term "non-specific catalytic activity" refers to the polypeptide(s) having an enzymatic activity capable of catalyzing the interconversion of 3-HPA and 1,3-propanediol and specifically excludes 1,3-propanediol oxidoreductase(s). Typically these enzymes are alcohol dehydrogenases. Such enzymes may utilize cofactors other than NAD⁺/NADH, including but not limited to flavins such as FAD or FMN. A gene for a non-specific alcohol dehydrogenase (*yqhD*) is found, for example, to be endogenously encoded and functionally expressed within *E. coli* K-12 strains.

The terms "1.6 long GI promoter", "1.20 short/long GI Promoter", and "1.5 long GI promoter" refer to polynucleotides or fragments containing a promoter from the *Streptomyces lividans* glucose isomerase gene as described in U.S. Patent No. 7,132,527. These promoter fragments include a mutation which decreases their activities as compared

to the wild type *Streptomyces lividans* glucose isomerase gene promoter.

The terms "function" and "enzyme function" are used interchangeably herein and refer to the catalytic activity of an enzyme in altering the rate at which a specific chemical reaction occurs without
5 itself being consumed by the reaction. It is understood that such an activity may apply to a reaction in equilibrium where the production of either product or substrate may be accomplished under suitable conditions.

The terms "polypeptide" and "protein" are used interchangeably
10 herein.

The terms "carbon substrate" and "carbon source" are used interchangeably herein and refer to a carbon source capable of being metabolized by the recombinant bacteria disclosed herein and, particularly, carbon sources comprising sucrose. The carbon source may
15 further comprise other monosaccharides, disaccharides, oligosaccharides; or polysaccharides.

The terms "host cell" and "host bacterium" are used interchangeably herein and refer to a bacterium capable of receiving foreign or heterologous genes and capable of expressing those genes to
20 produce an active gene product.

The term "production microorganism" as used herein refers to a microorganism, including, but not limited to, those that are recombinant, used to make a specific product such as 1,3-propanediol, glycerol, 3-hydroxypropionic acid, polyunsaturated fatty acids, and the like.

As used herein, "nucleic acid" means a polynucleotide and
25 includes a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases. Nucleic acids may also include fragments and modified nucleotides. Thus, the terms "polynucleotide", "nucleic acid sequence", "nucleotide sequence" or "nucleic acid fragment" are used
30 interchangeably herein and refer to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. Nucleotides (usually found in their 5'-

monophosphate form) are referred to by their single letter designation as follows: "A" for adenylate or deoxyadenylate (for RNA or DNA, respectively), "C" for cytidylate or deoxycytidylate, "G" for guanylate or deoxyguanylate, "U" for uridylate, "T" for deoxythymidylate, "R" for
5 purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, and "N" for any nucleotide.

A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may
10 be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, and which may refer to the coding region alone or may include regulatory sequences preceding (5' non-coding sequences) and
15 following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise
20 regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers
25 to a gene that is introduced into the host organism by gene transfer. Foreign genes can comprise genes inserted into a non-native organism, genes introduced into a new location within the native host, or chimeric genes.

The term "native nucleotide sequence" refers to a nucleotide
30 sequence that is normally found in the host microorganism.

The term "non-native nucleotide sequence" refers to a nucleotide sequence that is not normally found in the host microorganism.

The term "native polypeptide" refers to a polypeptide that is normally found in the host microorganism.

The term "non-native polypeptide" refers to a polypeptide that is not normally found in the host microorganism.

5 The terms "encoding" and "coding" are used interchangeably herein and refer to the process by which a gene, through the mechanisms of transcription and translation, produces an amino acid sequence.

The term "coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence.

10 "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters,
15 enhancers, silencers, 5' untranslated leader sequence (e.g., between the transcription start site and the translation initiation codon), introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

The term "expression cassette" refers to a fragment of DNA
20 comprising the coding sequence of a selected gene and regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence that are required for expression of the selected gene product. Thus, an expression cassette is typically composed of: 1) a promoter sequence; 2) a coding sequence (i.e., ORF)
25 and, 3) a 3' untranslated region (e.g., a terminator) that, in eukaryotes, usually contains a polyadenylation site. The expression cassette(s) is usually included within a vector, to facilitate cloning and transformation. Different organisms, including bacteria, yeast, and fungi, can be transformed with different expression cassettes as long as the correct
30 regulatory sequences are used for each host.

"Transformation" refers to the transfer of a nucleic acid molecule into a host organism, resulting in genetically stable inheritance. The

nucleic acid molecule may be a plasmid that replicates autonomously, for example, or it may integrate into the genome of the host organism. Host organisms transformed with the nucleic acid fragments are referred to as “recombinant” or “transformed” organisms or “transformants”. “Stable transformation” refers to the transfer of a nucleic acid fragment into a genome of a host organism, including both nuclear and organellar genomes, resulting in genetically stable inheritance. In contrast, “transient transformation” refers to the transfer of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance.

“Codon degeneracy” refers to the nature in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

The terms “subfragment that is functionally equivalent” and “functionally equivalent subfragment” are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment in which the ability to alter gene expression or produce a certain phenotype is retained whether or not the fragment or subfragment encodes an active enzyme. Chimeric genes can be designed for use in suppression by linking a nucleic acid fragment or subfragment thereof, whether or not it encodes an active enzyme, in the sense or antisense orientation relative to a promoter sequence.

The term “conserved domain” or “motif” means a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions indicate amino acids that are essential in the structure, the

stability, or the activity of a protein.

The terms "substantially similar" and "corresponds substantially" are used interchangeably herein. They refer to nucleic acid fragments wherein changes in one or more nucleotide bases do not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the invention encompasses more than the specific exemplary sequences. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize (under moderately stringent conditions, e.g., 0.5X SSC (standard sodium citrate), 0.1% SDS (sodium dodecyl sulfate), 60 °C) with the sequences exemplified herein, or to any portion of the nucleotide sequences disclosed herein and which are functionally equivalent to any of the nucleic acid sequences disclosed herein. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences are two nucleotide sequences wherein the complement of one of the nucleotide sequences typically has about at least 80% sequence identity, or 90% sequence identity, up to and including 100% sequence identity (i.e., fully

complementary) to the other nucleotide sequence.

The term "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will selectively hybridize to its target sequence. Probes are typically single
5 stranded nucleic acid sequences which are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Hybridization methods are well defined. Typically the probe and
10 sample are mixed under conditions which will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. Optionally a chaotropic agent may be added. Nucleic acid hybridization is adaptable to a variety of assay formats. One
15 of the most suitable is the sandwich assay format. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it an immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Stringent conditions are sequence-dependent and will be different
20 in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are
25 detected (heterologous probing).

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes (e.g., 10 to 50 nucleotides) and at
30 least about 60 °C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency

conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37 °C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55 °C. Exemplary moderate stringency conditions include

5 hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37 °C, and a wash in 0.5X to 1X SSC at 55 to 60 °C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37 °C, and a wash in 0.1X SSC at 60 to 65 °C.

Specificity is typically the function of post-hybridization washes, the

10 critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the thermal melting point (T_m) can be approximated from the equation of Meinkoth et al., *Anal. Biochem.* 138:267-284 (1984): $T_m = 81.5\text{ °C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, %GC is the

15 percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by

20 about 1 °C for each 1% of mismatching; thus, T_m , hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq 90\%$ identity are sought, the T_m can be decreased 10 °C. Generally, stringent conditions are selected to be about 5 °C lower than T_m for the specific sequence and its complement

25 at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4 °C lower than the T_m ; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the T_m ; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20 °C lower than the

30 T_m . Using the equation, hybridization and wash compositions, and desired T_m , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the

desired degree of mismatching results in a T_m of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory*
5 *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel et al., Eds., Greene Publishing and Wiley-Interscience, New York
10 (1995). Hybridization and/or wash conditions can be applied for at least 10, 30, 60, 90, 120, or 240 minutes.

"Sequence identity" or "identity" in the context of nucleic acid or polypeptide sequences refers to the nucleic acid bases or amino acid residues in two sequences that are the same when aligned for maximum
15 correspondence over a specified comparison window.

Thus, "percentage of sequence identity" refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions
20 or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions,
25 dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 50% to 100%.
30 These identities can be determined using any of the programs described herein.

Sequence alignments and percent identity or similarity calculations

may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the “default values” of the program referenced, unless otherwise specified. As used herein “default values” will mean any set of values or parameters that originally load with the software when first initialized.

10 The “Clustal V method of alignment” corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp, *CABIOS* 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.* 8:189-191 (1992)) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal V method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are

15 KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program.

25 The “Clustal W method of alignment” corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, *supra*; Higgins, D.G. et al., *supra*) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for multiple alignment correspond to GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen

30 Seqs(%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB. After alignment of the sequences using the Clustal W program, it is possible to obtain a “percent identity” by

viewing the "sequence distances" table in the same program.

"BLASTN method of alignment" is an algorithm provided by the National Center for Biotechnology Information (NCBI) to compare nucleotide sequences using default parameters.

5 It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any
10 integer percentage from 50% to 100%. Indeed, any integer amino acid identity from 50% to 100% may be useful in describing the present invention, such as 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%,
15 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Also, of interest is any full-length or partial complement of this isolated nucleotide fragment.

Thus, the invention encompasses more than the specific exemplary nucleotide sequences disclosed herein. For example, alterations in the
20 gene sequence which reflect the degeneracy of the genetic code are contemplated. Also, it is well known in the art that alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not affect the functional properties of the encoded protein are common. Substitutions are defined for the discussion herein
25 as exchanges within one of the following five groups:

1. Small aliphatic, nonpolar or slightly polar residues: Ala, Ser, Thr (Pro, Gly);
2. Polar, negatively charged residues and their amides: Asp, Asn, Glu, Gln;
- 30 3. Polar, positively charged residues: His, Arg, Lys;
4. Large aliphatic, nonpolar residues: Met, Leu, Ile, Val (Cys); and

5. Large aromatic residues: Phe, Tyr, Trp.

Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue (such as glycine) or a more hydrophobic residue (such as valine, leucine, or isoleucine). Similarly, changes which result in substitution of one negatively charged residue for another (such as aspartic acid for glutamic acid) or one positively charged residue for another (such as lysine for arginine) can also be expected to produce a functionally equivalent product. In many cases, nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein.

Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their ability to hybridize under stringent conditions, as defined above.

Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose nucleotide sequences are at least 70% identical to the nucleotide sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are at least 90% identical to the nucleotide sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are at least 95% identical to the nucleotide sequence of the nucleic acid fragments reported herein.

A "substantial portion" of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.*, 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or

more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be
5 used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers.
10 Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The instant specification teaches the complete amino acid and nucleotide sequence encoding particular proteins. The skilled artisan, having the benefit of the sequences as
15 reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art.

The term "complementary" describes the relationship between two sequences of nucleotide bases that are capable of Watson-Crick base-pairing when aligned in an anti-parallel orientation. For example, with
20 respect to DNA, adenosine is capable of base-pairing with thymine and cytosine is capable of base-pairing with guanine. Accordingly, the instant invention may make use of isolated nucleic acid molecules that are complementary to the complete sequences as reported in the accompanying Sequence Listing and the specification as well as those
25 substantially similar nucleic acid sequences.

The term "isolated" refers to a polypeptide or nucleotide sequence that is removed from at least one component with which it is naturally associated.

"Promoter" refers to a DNA sequence capable of controlling the
30 expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer"

is a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters".

"3' non-coding sequences", "transcription terminator" and "termination sequences" are used interchangeably herein and refer to DNA sequences located downstream of a coding sequence, including polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation. In another example, the complementary RNA regions of the invention can be operably linked, either directly or indirectly, 5' to the target mRNA, or 3' to the target mRNA, or within the target mRNA, or a first complementary

region is 5' and its complement is 3' to the target mRNA.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; 5 Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989). Transformation methods are well known to those skilled in the art and are described *infra*.

"PCR" or "polymerase chain reaction" is a technique for the synthesis of large quantities of specific DNA segments and consists of a series of repetitive cycles (Perkin Elmer Cetus Instruments, Norwalk, CT). 10 Typically, the double-stranded DNA is heat denatured, the two primers complementary to the 3' boundaries of the target segment are annealed at low temperature and then extended at an intermediate temperature. One set of these three consecutive steps is referred to as a "cycle".

15 A "plasmid" or "vector" is an extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, 20 of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing an expression cassette(s) into a cell.

The term "genetically altered" refers to the process of changing 25 hereditary material by genetic engineering, transformation and/or mutation.

The term "recombinant" refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic 30 engineering techniques. "Recombinant" also includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or a cell derived from a cell so modified, but does not

encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation, natural transduction, natural transposition) such as those occurring without deliberate human intervention.

5 The terms “recombinant construct”, “expression construct”, “chimeric construct”, “construct”, and “recombinant DNA construct”, are used interchangeably herein. A recombinant construct comprises an artificial combination of nucleic acid fragments, e.g., regulatory and coding sequences that are not found together in nature. For example, a
10 recombinant construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Such a construct may be used by itself or may be used in conjunction with a vector. If a
15 vector is used, then the choice of vector is dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells
20 comprising any of the isolated nucleic acid fragments of the invention. The skilled artisan will also recognize that different independent transformation events may result in different levels and patterns of expression (Jones et al., *EMBO J.* 4:2411-2418 (1985); De Almeida et al., *Mol. Gen. Genetics* 218:78-86 (1989)), and thus that multiple events may
25 need be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others.

30 The term “expression”, as used herein, refers to the production of a functional end-product (e.g., an mRNA or a protein [either precursor or mature]).

The term "introduced" means providing a nucleic acid (e.g., expression construct) or protein into a cell. Introduced includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell, and includes reference to the transient provision of a nucleic acid or protein to the cell. Introduced includes reference to stable or transient transformation methods, as well as sexually crossing. Thus, "introduced" in the context of inserting a nucleic acid fragment (e.g., a recombinant construct/expression construct) into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid fragment into a eukaryotic or prokaryotic cell where the nucleic acid fragment may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The term "homologous" refers to proteins or polypeptides of common evolutionary origin with similar catalytic function. The invention may include bacteria producing homologous proteins via recombinant technology.

Disclosed herein are recombinant bacteria that have been engineered to utilize sucrose using new sucrose utilization genes from *Pseudomonas fluorescens* (*P. fluorescens*). Specifically, the recombinant bacteria disclosed herein comprise in their genome or on at least one recombinant construct: a nucleotide sequence encoding a polypeptide having sucrose transporter activity, the polypeptide having at least 95% sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:24; and a nucleotide sequence encoding a polypeptide having sucrose hydrolase activity, the polypeptide having at least 95% sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:26. The nucleotide sequences are each operably linked to the same or a different promoter. In some

embodiments, the recombinant bacteria are capable of metabolizing sucrose to produce glycerol and/or glycerol-derived products.

Suitable host bacteria for use in the construction of the recombinant bacteria disclosed herein include, but are not limited to, organisms of the
5 genera: *Escherichia*, *Streptococcus*, *Agrobacterium*, *Bacillus*,
Corynebacterium, *Lactobacillus*, *Clostridium*, *Gluconobacter*, *Citrobacter*,
Enterobacter, *Klebsiella*, *Aerobacter*, *Methylobacter*, *Salmonella*,
Streptomyces, and *Pseudomonas*.

In one embodiment, the host bacterium is selected from the genera:
10 *Escherichia*, *Klebsiella*, *Citrobacter*, and *Aerobacter*.

In another embodiment, the host bacterium is *Escherichia coli*.

In some embodiments, the host bacterium is PTS minus. In these
embodiments, the host bacterium is PTS minus in its native state, or may
be rendered PTS minus through inactivation of a PTS gene as described
15 below.

In production microorganisms, it is sometimes desirable to unlink
the transport of sugars and the use of phosphoenolpyruvate (PEP) for
phosphorylation of the sugars being transported.

The term "down-regulated" refers to reduction in, or abolishment of,
20 the activity of active protein(s), as compared to the activity of the wildtype
protein(s). The PTS may be inactivated (resulting in a "PTS minus"
organism) by down-regulating expression of one or more of the
endogenous genes encoding the proteins required in this type of transport.
Down-regulation typically occurs when one or more of these genes has a
25 "disruption", referring to an insertion, deletion, or targeted mutation within
a portion of that gene, that results in either a complete gene knockout
such that the gene is deleted from the genome and no protein is translated
or a protein has been translated such that it has an insertion, deletion,
amino acid substitution or other targeted mutation. The location of the
30 disruption in the protein may be, for example, within the N-terminal portion
of the protein or within the C-terminal portion of the protein. The disrupted
protein will have impaired activity with respect to the protein that was not

disrupted, and can be non-functional. Down-regulation that results in low or lack of expression of the protein, could also result via manipulating the regulatory sequences, transcription and translation factors and/or signal transduction pathways or by use of sense, antisense or RNAi technology, etc.

Sucrose transporter polypeptides are polypeptides that are capable of mediating the transport of sucrose into microbial cells. Sucrose transport polypeptides are known in the art, for example the CscB polypeptide from *E. coli* wild-type strain EC3132 (set forth in SEQ ID NO:28), encoded by the *cscB* gene (coding sequence set forth in SEQ ID NO:27), as described by Jahreis et al. (*J. Bacteriol.* 184:5307-5316, 2002). The *lacY* gene from *Pseudomonas fluorescens* Pf5 ATCC® BAA-477 (coding sequence set forth in SEQ ID NO:23) encodes a polypeptide (set forth in SEQ ID NO:24) that is 41% identical to the known *Escherichia coli* strain EC3132 sucrose transport protein, CscB. However, there is no known activity demonstrated for this polypeptide.

In one embodiment, the nucleotide sequence encoding a polypeptide having sucrose transporter activity has at least 95% sequence identity, based on BLASTN method of alignment, when compared to a nucleotide sequence as set forth in SEQ ID NO:23.

In another embodiment, the nucleotide sequence encoding a polypeptide having sucrose transporter activity has the nucleotide sequence set forth in SEQ ID NO:23.

In one embodiment, the polypeptide having sucrose transporter activity has at least 95% sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:24.

In another embodiment, the polypeptide having sucrose transporter activity has the amino acid sequence set forth in SEQ ID NO:24.

Polypeptides having sucrose hydrolase activity have the ability to catalyze the hydrolysis of sucrose to produce fructose and glucose. Polypeptides having sucrose hydrolase activity are known, and include,

CscA from *E. coli* wild-type strain EC3132 (set forth in SEQ ID NO:30), encoded by gene *cscA* (coding sequence set forth in SEQ ID NO:29), as described by Jahreis et al. (*J. Bacteriol.* 184:5307-5316, 2002). The *scrB* gene from *Pseudomonas fluorescens* Pf5 ATCC® BAA-477 (coding
5 sequence set forth in SEQ ID NO:25) encodes a polypeptide (set forth in SEQ ID NO:26) that is 47% identical to the known *Escherichia coli* strain EC3132 sucrose hydrolase, CscA. However, there is no known activity demonstrated for this polypeptide.

In one embodiment, the nucleotide sequence encoding a
10 polypeptide having sucrose hydrolase activity has at least 95% sequence identity, based on BLASTN method of alignment, when compared to a nucleotide sequence as set forth in SEQ ID NO:25.

In another embodiment, the nucleotide sequence encoding a polypeptide having sucrose hydrolase activity has the nucleotide
15 sequence set forth in SEQ ID NO:25.

In one embodiment, the polypeptide having sucrose hydrolase activity has at least 95% sequence identity, based on a Clustal V method of alignment, when compared to an amino acid sequence as set forth in SEQ ID NO:26.

20 In another embodiment, the polypeptide having sucrose hydrolase activity has the amino acid sequence set forth in SEQ ID NO:26.

The recombinant bacteria disclosed herein may further comprise in their genome or on at least one recombinant construct, a nucleotide sequence encoding a polypeptide having fructokinase activity to enable
25 the bacteria to utilize the fructose produced by the hydrolysis of sucrose. Polypeptides having fructokinase activity include fructokinases (designated EC 2.7.1.4) and various hexose kinases having fructose phosphorylating activity (EC 2.7.1.3 and EC 2.7.1.1). Fructose phosphorylating activity may be exhibited by hexokinases and
30 ketohexokinases. Representative genes encoding polypeptides from a variety of microorganisms, which may be used to construct the recombinant bacteria disclosed herein, are listed in Table 1. One skilled in

the art will know that proteins that are substantially similar to a protein which is able to phosphorylate fructose (such as encoded by the genes listed in Table 1) may also be used.

5

Table 1

Sequences Encoding Enzymes with Fructokinase Activity

Source	Gene Name	EC Number	Nucleotide SEQ ID NO:	Protein SEQ ID NO:
<i>Agrobacterium tumefaciens</i>	<i>scrK</i> (fructokinase)	2.7.1.4	31	32
<i>Streptococcus mutans</i>	<i>scrK</i> (fructokinase)	2.7.1.4	33	34
<i>Escherichia coli</i>	<i>scrK</i> (fructokinase)	2.7.1.4	35	36
<i>Klebsiella pneumoniae</i>	<i>scrK</i> (fructokinase)	2.7.1.4	37	38
<i>Escherichia coli</i>	<i>cscK</i> (fructokinase)	2.7.1.4	39	40
<i>Enterococcus faecalis</i>	<i>cscK</i> (fructokinase)	2.7.1.4	41	42
<i>Saccharomyces cerevisiae</i>	<i>HXK1</i> (hexokinase)	2.7.1.1	43	44
<i>Saccharomyces cerevisiae</i>	<i>HXK2</i> (hexokinase)	2.7.1.1	45	46

In one embodiment, the polypeptide having fructokinase activity has at least 95% sequence identity, based on the Clustal V method of alignment, to an amino acid sequence as set forth in SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, or SEQ ID NO:46.

In another embodiment, the polypeptide having fructokinase activity has the amino acid sequence set forth in SEQ ID NO:40.

The coding sequence of the genes encoding polypeptides having sucrose transporter activity and polypeptides having sucrose hydrolase activity may be used to isolate nucleotide sequences encoding homologous polypeptides from the same or other microbial species. For example, homologs of the genes may be identified using sequence analysis software, such as BLASTN, to search publically available nucleic

acid sequence databases. Additionally, the isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g. polymerase chain reaction (PCR), Mullis et al., U.S. Patent No. 4,683,202; ligase chain reaction (LCR), Tabor, S. et al., *Proc. Acad. Sci. USA* 82, 1074, 1985); or strand displacement amplification (SDA), Walker, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89: 392, (1992)). For example, the nucleotide sequence encoding the polypeptides described above may be employed as a hybridization probe for the identification of homologs.

One of ordinary skill in the art will appreciate that genes encoding these polypeptides isolated from other sources may also be used in the recombinant bacteria disclosed herein. Additionally, variations in the nucleotide sequences encoding the polypeptides may be made without affecting the amino acid sequence of the encoded polypeptide due to codon degeneracy, and that amino acid substitutions, deletions or additions that produce a substantially similar protein may be included in the encoded protein.

The nucleotide sequences encoding the polypeptides having sucrose transporter activity and polypeptides having sucrose hydrolase activity may be isolated using PCR (see, e.g., U.S. Patent No. 4,683,202) and primers designed to bound the desired sequence, for example as described in Example 1 herein. Other methods of gene isolation are well known to one skilled in the art such as by using degenerate primers or heterologous probe hybridization. The nucleotide sequences can also be chemically synthesized or purchased from vendors such as DNA2.0 Inc. (Menlo Park, CA).

Expression of the polypeptides may be effected using one of many methods known to one skilled in the art. For example, the nucleotide sequences encoding the polypeptides described above may be introduced into the bacterium on at least one multicopy plasmid, or by

integrating one or more copies of the coding sequences into the host genome. The nucleotide sequences encoding the polypeptides may be introduced into the host bacterium separately (e.g., on separate plasmids) or in any combination (e.g., on a single plasmid, as described
5 in the Examples herein).

The introduced coding regions that are either on a plasmid(s) or in the genome may be expressed from at least one highly active promoter. An integrated coding region may either be introduced as a part of a chimeric gene having its own promoter, or it may be integrated
10 adjacent to a highly active promoter that is endogenous to the genome or in a highly expressed operon. Suitable promoters include, but are not limited to, *CYC1*, *HIS3*, *GAL1*, *GAL10*, *ADH1*, *PGK*, *PHO5*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO*, and *lac*, *ara*, *tet*, *trp*, *IP_L*, *IP_R*, *T7*, *tac*, and *trc* (useful for expression in *Escherichia coli*) as well as the *amy*,
15 *apr*, *npr* promoters and various phage promoters useful for expression in *Bacillus*. The promoter may also be the *Streptomyces lividans* glucose isomerase promoter or a variant thereof, described by Payne et al. (U.S. Patent No. 7,132,527).

In one embodiment, the recombinant bacteria disclosed herein
20 are capable of producing glycerol. Biological processes for the preparation of glycerol using carbohydrates or sugars are known in yeasts and in some bacteria, other fungi, and algae. Both bacteria and yeasts produce glycerol by converting glucose or other carbohydrates through the fructose-1,6-bisphosphate pathway in glycolysis. In the
25 method of producing glycerol disclosed herein, host bacteria may be used that naturally produce glycerol. In addition, bacteria may be engineered for production of glycerol and glycerol derivatives. The capacity for glycerol production from a variety of substrates may be provided through the expression of the enzyme activities glycerol-3-
30 phosphate dehydrogenase (G3PDH) and/or glycerol-3-phosphatase as described in U.S. Patent No. 7,005,291. Genes encoding these proteins that may be used for expressing the enzyme activities in a host

bacterium are described in U.S. Patent No. 7,005,291. Suitable examples of genes encoding polypeptides having glycerol-3-phosphate dehydrogenase activity include, but are not limited to, *GPD1* from *Saccharomyces cerevisiae* (coding sequence set forth in SEQ ID NO:1, encoded protein sequence set forth in SEQ ID NO:2) and *GPD2* from *Saccharomyces cerevisiae* (coding sequence set forth in SEQ ID NO:3, encoded protein sequence set forth in SEQ ID NO:4). Suitable examples of genes encoding polypeptides having glycerol-3-phosphatase activity include, but are not limited to, *GPP1* from *Saccharomyces cerevisiae* (coding sequence set forth in SEQ ID NO:5, encoded protein sequence set forth in SEQ ID NO:6) and *GPP2* from *Saccharomyces cerevisiae* (coding sequence set forth in SEQ ID NO:7, encoded protein sequence set forth in SEQ ID NO:8).

Increased production of glycerol may be attained through reducing expression of target endogenous genes. Down-regulation of endogenous genes encoding glycerol kinase and glycerol dehydrogenase activities further enhance glycerol production as described in U.S. Patent No. 7,005,291. Increased channeling of carbon to glycerol may be accomplished by reducing the expression of the endogenous gene encoding glyceraldehyde 3-phosphate dehydrogenase, as described in U.S. Patent No. 7,371,558. Down-regulation may be accomplished by using any method known in the art, for example, the methods described above for down-regulation of genes of the PTS system.

Glycerol provides a substrate for microbial production of useful products. Examples of such products, i.e., glycerol derivatives include, but are not limited to, 3-hydroxypropionic acid, methylglyoxal, 1,2-propanediol, and 1,3-propanediol.

In another embodiment, the recombinant bacteria disclosed herein are capable of producing 1,3-propanediol. The glycerol derivative 1,3-propanediol is a monomer having potential utility in the production of polyester fibers and the manufacture of polyurethanes and cyclic compounds. 1,3-Propanediol can be produced by a single microorganism

by bioconversion of a carbon substrate other than glycerol or dihydroxyacetone, as described in U.S. Patent No. 5,686,276. In this bioconversion, glycerol is produced from the carbon substrate, as described above. Glycerol is converted to the intermediate 3-hydroxypropionaldehyde by a dehydratase enzyme, which can be encoded by the host bacterium or can be introduced into the host by recombination. The dehydratase can be glycerol dehydratase (E.C. 4.2.1.30), diol dehydratase (E.C. 4.2.1.28) or any other enzyme able to catalyze this conversion. A suitable example of genes encoding the “ α ” (alpha), “ β ” (beta), and “ γ ” (gamma) subunits of a glycerol dehydratase include, but are not limited to *dhaB1* (coding sequence set forth in SEQ ID NO:9), *dhaB2* (coding sequence set forth in SEQ ID NO:11), and *dhaB3* (coding sequence set forth in SEQ ID NO:13), respectively, from *Klebsiella pneumoniae*. The further conversion of 3-hydroxypropionaldehyde to 1,3-propanediol can be catalyzed by 1,3-propanediol dehydrogenase (E.C. 1.1.1.202) or other alcohol dehydrogenases. A suitable example of a gene encoding a 1,3-propanediol dehydrogenase is *dhaT* from *Klebsiella pneumoniae* (coding sequence set forth in SEQ ID NO:47, encoded protein sequence set forth in SEQ ID NO:48).

Bacteria can be recombinantly engineered to provide more efficient production of glycerol and the glycerol derivative 1,3-propanediol. For example, U.S. Patent No. 7,005,291 discloses transformed microorganisms and a method for production of glycerol and 1,3-propanediol with advantages derived from expressing exogenous activities of one or both of glycerol-3-phosphate dehydrogenase and glycerol-3-phosphate phosphatase while disrupting one or both of endogenous activities glycerol kinase and glycerol dehydrogenase.

U.S. Patent No. 6,013,494 describes a process for the production of 1,3-propanediol using a single microorganism comprising exogenous glycerol-3-phosphate dehydrogenase, glycerol-3-phosphate phosphatase, dehydratase, and 1,3-propanediol oxidoreductase (e.g., *dhaT*). U.S. Patent No. 6,136,576 discloses a method for the production of 1,3-

propanediol comprising a recombinant microorganism further comprising a dehydratase and protein X (later identified as being a dehydratase reactivation factor peptide).

U.S. Patent No. 6,514,733 describes an improvement to the process where a significant increase in titer (grams product per liter) is obtained by virtue of a non-specific catalytic activity (distinguished from 1,3-propanediol oxidoreductase encoded by *dhaT*) to convert 3-hydroxypropionaldehyde to 1,3-propanediol. Additionally, U.S. Patent No. 7,132,527 discloses vectors and plasmids useful for the production of 1,3-propanediol.

Increased production of 1,3-propanediol may be achieved by further modifications to a host bacterium, including down-regulating expression of some target genes and up-regulating, expression of other target genes, as described in U.S. Patent No. 7,371,558. For utilization of glucose as a carbon source in a PTS minus host, expression of glucokinase activity may be increased.

Additional genes whose increased or up-regulated expression increases 1,3-propanediol production include genes encoding:

- phosphoenolpyruvate carboxylase typically characterized as EC 4.1.1.31
- cob(I)alamin adenosyltransferase, typically characterized as EC 2.5.1.17
- non-specific catalytic activity that is sufficient to catalyze the interconversion of 3-HPA and 1,3-propanediol, and specifically excludes 1,3-propanediol oxidoreductase(s), typically these enzymes are alcohol dehydrogenases

Genes whose reduced or down-regulated expression increases 1,3-propanediol production include genes encoding:

- aerobic respiration control protein
- methylglyoxal synthase
- acetate kinase
- phosphotransacetylase

- aldehyde dehydrogenase A
- aldehyde dehydrogenase B
- triosephosphate isomerase
- phosphogluconate dehydratase

5 In another embodiment, the recombinant bacteria disclosed herein are capable of producing 3-hydroxypropionic acid. 3-Hydroxypropionic acid has utility for specialty synthesis and can be converted to commercially important intermediates by known art in the chemical industry, e.g., acrylic acid by dehydration, malonic acid by oxidation,
10 esters by esterification reactions with alcohols, and 1,3-propanediol by reduction. 3-Hydroxypropionic acid may be produced biologically from a fermentable carbon source by a single microorganism, as described in copending and commonly owned U.S. Patent Application No. 12/815461. In one representative biosynthetic pathway, a carbon substrate is
15 converted to 3-hydroxypropionaldehyde, as described above for the production of 1,3-propanediol. The 3-hydroxypropionaldehyde is converted to 3-hydroxypropionic acid by an aldehyde dehydrogenase. Suitable examples of aldehyde dehydrogenases include, but are not limited to, AldB (SEQ ID NO:16), encoded by the *E. coli* gene *aldB* (coding
20 sequence set forth in SEQ ID NO:15); AldA (SEQ ID NO:18), encoded by the *E. coli* gene *aldA* (coding sequence set forth in SEQ ID NO:17); and AldH (SEQ ID NO:20), encoded by the *E. coli* gene *aldH* (coding sequence asset forth in SEQ ID NO:19).

 Many of the modifications described above to improve 1,3-
25 propanediol production by a recombinant bacterium can also be made to improve 3-hydroxypropionic acid production. For example, the elimination of glycerol kinase prevents glycerol, formed from G3P by the action of G3P phosphatase, from being re-converted to G3P at the expense of ATP. Also, the elimination of glycerol dehydrogenase (for example, *gldA*)
30 prevents glycerol, formed from DHAP by the action of NAD-dependent glycerol-3-phosphate dehydrogenase, from being converted to dihydroxyacetone. Mutations can be directed toward a structural gene so

as to impair or improve the activity of an enzymatic activity or can be directed toward a regulatory gene, including promoter regions and ribosome binding sites, so as to modulate the expression level of an enzymatic activity.

5 Up-regulation or down-regulation may be achieved by a variety of methods which are known to those skilled in the art. It is well understood that up-regulation or down-regulation of a gene refers to an alteration in the level of activity present in a cell that is derived from the protein encoded by that gene relative to a control level of activity, for example, by
10 the activity of the protein encoded by the corresponding (or non-altered) wild-type gene.

Specific genes involved in an enzyme pathway may be up-regulated to increase the activity of their encoded function(s). For example, additional copies of selected genes may be introduced into the
15 host cell on multicopy plasmids such as pBR322. Such genes may also be integrated into the chromosome with appropriate regulatory sequences that result in increased activity of their encoded functions. The target genes may be modified so as to be under the control of non-native promoters or altered native promoters. Endogenous promoters
20 can be altered *in vivo* by mutation, deletion, and/or substitution.

Alternatively, it may be useful to reduce or eliminate the expression of certain genes relative to a given activity level. Methods of down-regulating (disrupting) genes are known to those of skill in the art.

Down-regulation can occur by deletion, insertion, or alteration of
25 coding regions and/or regulatory (promoter) regions. Specific down regulations may be obtained by random mutation followed by screening or selection, or, where the gene sequence is known, by direct intervention by molecular biology methods known to those skilled in the art. A particularly useful, but not exclusive, method to effect down-regulation is to alter
30 promoter strength.

Furthermore, down-regulation of gene expression may be used to either prevent expression of the protein of interest or result in the

expression of a protein that is non-functional. This may be accomplished for example, by 1) deleting coding regions and/or regulatory (promoter) regions, 2) inserting exogenous nucleic acid sequences into coding regions and/or regulatory (promoter) regions, and 3) altering coding regions and/or regulatory (promoter) regions (for example, by making DNA base pair changes). Specific disruptions may be obtained by random mutation followed by screening or selection, or, in cases where the gene sequences in known, specific disruptions may be obtained by direct intervention using molecular biology methods known to those skilled in the art. A particularly useful method is the deletion of significant amounts of coding regions and/or regulatory (promoter) regions.

Methods of altering recombinant protein expression are known to those skilled in the art, and are discussed in part in Baneyx, *Curr. Opin. Biotechnol.* (1999) 10:411; Ross, et al., *J. Bacteriol.* (1998) 180:5375; deHaseth, et al., *J. Bacteriol.* (1998) 180:3019; Smolke and Keasling, *Biotechnol. Bioeng.* (2002) 80:762; Swartz, *Curr. Opin. Biotech.* (2001) 12:195; and Ma, et al., *J. Bacteriol.* (2002) 184:5733.

Recombinant bacteria containing the necessary changes in gene expression for metabolizing sucrose in the production of microbial products including glycerol and glycerol derivatives, as described above, may be constructed using techniques well known in the art, some of which are exemplified in the Examples herein.

The construction of the recombinant bacteria disclosed herein may be accomplished using a variety of vectors and transformation and expression cassettes suitable for the cloning, transformation and expression of coding regions that confer the ability to utilize sucrose in the production of glycerol and its derivatives in a suitable host microorganism. Suitable vectors are those which are compatible with the bacterium employed. Suitable vectors can be derived, for example, from a bacterium, a virus (such as bacteriophage T7 or a M-13 derived phage), a cosmid, a yeast or a plant. Protocols for obtaining and using such vectors are known to those skilled in the art (Sambrook *et al.*, *supra*).

Initiation control regions, or promoters, which are useful to drive expression of coding regions for the instant invention in the desired host bacterium are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving expression is suitable for use herein. For example, any of the promoters listed above may be used.

Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

For effective expression of the instant polypeptides, nucleotide sequences encoding the polypeptides are linked operably through initiation codons to selected expression control regions such that expression results in the formation of the appropriate messenger RNA.

Particularly useful are the vectors pSYCO101, pSYCO103, pSYCO106, and pSYCO109, described in U.S. Patent No. 7,371,558, and pSYCO400/AGRO, described in U.S. Patent No. 7,524,660. The essential elements of these vectors are derived from the *dha* regulon isolated from *Klebsiella pneumoniae* and from *Saccharomyces cerevisiae*. Each vector contains the open reading frames *dhaB1*, *dhaB2*, *dhaB3*, *dhaX* (coding sequence set forth in SEQ ID NO:49; encoded polypeptide sequence set forth in SEQ ID NO:50), *orfX*, *DAR1*, and *GPP2* arranged in three separate operons. The nucleotide sequences of pSYCO101, pSYCO103, pSYCO106, pSYCO109, and pSYCO400/AGRO are set forth in SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, and SEQ ID NO:55, respectively. The differences between the vectors are illustrated in the chart below [the prefix "p-" indicates a promoter; the open reading frames contained within each "()" represent the composition of an operon]:

pSYCO101 (SEQ ID NO:51):

p-trc (*Dar1_GPP2*) in opposite orientation compared to the other 2 pathway operons,
p-1.6 long GI (*dhaB1_dhaB2_dhaB3_dhaX*), and
p-1.6 long GI (*orfY_orfX_orfW*).

pSYCO103 (SEQ ID NO:52):

p-trc (Dar1_GPP2) same orientation compared to the other 2
pathway operons,

p-1.5 long GI (dhaB1_dhaB2_dhaB3_dhaX), and

5 p-1.5 long GI (orfY_orfX_orfW).

pSYCO106 (SEQ ID NO:53):

p-trc (Dar1_GPP2) same orientation compared to the other 2
pathway operons,

p-1.6 long GI (dhaB1_dhaB2_dhaB3_dhaX), and

10 p-1.6 long GI (orfY_orfX_orfW).

pSYCO109 (SEQ ID NO:54):

p-trc (Dar1_GPP2) same orientation compared to the other 2
pathway operons,

p-1.6 long GI (dhaB1_dhaB2_dhaB3_dhaX), and

15 p-1.6 long GI (orfY_orfX).

pSYCO400/AGRO (SEQ ID NO:55):

p-trc (Dar1_GPP2) same orientation compared to the other 2
pathway operons,

p-1.6 long GI (dhaB1_dhaB2_dhaB3_dhaX), and

20 p-1.6 long GI (orfY_orfX).

p-1.20 short/long GI (scrK) opposite orientation compared to the
pathway operons.

Once suitable expression cassettes are constructed, they are used
to transform appropriate host bacteria. Introduction of the cassette
25 containing the coding regions into the host bacterium may be
accomplished by known procedures such as by transformation (e.g., using
calcium-permeabilized cells, or electroporation) or by transfection using a
recombinant phage virus (Sambrook et al., *supra*). Expression cassettes
may be maintained on a stable plasmid in a host cell. In addition,
30 expression cassettes may be integrated into the genome of the host
bacterium through homologous or random recombination using vectors
and methods well known to those skilled in the art. Site-specific

recombination systems may also be used for genomic integration of expression cassettes.

In addition to the cells exemplified, cells having single or multiple mutations specifically designed to enhance the production of microbial products including glycerol and/or its derivatives may also be used. Cells that normally divert a carbon feed stock into non-productive pathways, or that exhibit significant catabolite repression may be mutated to avoid these phenotypic deficiencies.

Methods of creating mutants are common and well known in the art. A summary of some methods is presented in U.S. Patent No. 7,371,558. Specific methods for creating mutants using radiation or chemical agents are well documented in the art. See, for example, Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA., or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.* 36, 227 (1992).

After mutagenesis has occurred, mutants having the desired phenotype may be selected by a variety of methods. Random screening is most common where the mutagenized cells are selected for the ability to produce the desired product or intermediate. Alternatively, selective isolation of mutants can be performed by growing a mutagenized population on selective media where only resistant colonies can develop. Methods of mutant selection are highly developed and well known in the art of industrial microbiology. See, for example, Brock, *Supra*; DeMancilha *et al.*, *Food Chem.* 14, 313 (1984).

Fermentation media in the present invention comprise sucrose as a carbon substrate. Other carbon substrates such as glucose and fructose may also be present.

In addition to the carbon substrate, a suitable fermentation medium contains, for example, suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures and promotion of the enzymatic pathway necessary for production of glycerol and its derivatives, for example 1,3-propanediol.

Particular attention is given to Co(II) salts and/or vitamin B₁₂ or precursors thereof in production of 1,3-propanediol.

Adenosyl-cobalamin (coenzyme B₁₂) is an important cofactor for dehydratase activity. Synthesis of coenzyme B₁₂ is found in prokaryotes, some of which are able to synthesize the compound *de novo*, for example, *Escherichia blattae*, *Klebsiella* species, *Citrobacter* species, and *Clostridium* species, while others can perform partial reactions. *E. coli*, for example, cannot fabricate the corrin ring structure, but is able to catalyze the conversion of cobinamide to corrinoid and can introduce the 5'-deoxyadenosyl group. Thus, it is known in the art that a coenzyme B₁₂ precursor, such as vitamin B₁₂, needs be provided in *E. coli* fermentations. Vitamin B₁₂ may be added continuously to *E. coli* fermentations at a constant rate or staged as to coincide with the generation of cell mass, or may be added in single or multiple bolus additions.

Although vitamin B₁₂ is added to the transformed *E. coli* described herein, it is contemplated that other bacteria, capable of *de novo* vitamin B₁₂ biosynthesis will also be suitable production cells and the addition of vitamin B₁₂ to these bacteria will be unnecessary.

Typically bacterial cells are grown at 25 to 40 °C in an appropriate medium containing sucrose. Examples of suitable growth media for use herein are common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used, and the appropriate medium for growth of the particular bacterium will be known by someone skilled in the art of microbiology or fermentation science. The use of agents known to modulate catabolite repression directly or indirectly, e.g., cyclic adenosine 2':3'-monophosphate, may also be incorporated into the reaction media. Similarly, the use of agents known to modulate enzymatic activities (e.g., methyl viologen) that lead to enhancement of 1,3-propanediol production may be used in conjunction

with or as an alternative to genetic manipulations with 1,3-propanediol production strains.

Suitable pH ranges for the fermentation are between pH 5.0 to pH 9.0, where pH 6.0 to pH 8.0 is typical as the initial condition.

5 Reactions may be performed under aerobic, anoxic, or anaerobic conditions depending on the requirements of the recombinant bacterium. Fed-batch fermentations may be performed with carbon feed, for example, carbon substrate, limited or excess.

10 Batch fermentation is a commonly used method. Classical batch fermentation is a closed system where the composition of the medium is set at the beginning of the fermentation and is not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation, the medium is inoculated with the desired bacterium and fermentation is permitted to occur adding nothing to the system.

15 Typically, however, "batch" fermentation is batch with respect to the addition of carbon source, and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems, the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures, cells

20 moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

25 A variation on the standard batch system is the Fed-Batch system. Fed-Batch fermentation processes are also suitable for use herein and comprise a typical batch system with the exception that the substrate is added in increments as the fermentation progresses. Fed-Batch systems are useful when catabolite repression is apt to inhibit the metabolism of

30 the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of

the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂. Batch and Fed-Batch fermentations are common and well known in the art and examples may be found in Brock, *supra*.

5 Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

10 Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems, a number of factors affecting
15 growth can be altered continuously while the cell concentration, measured by the turbidity of the medium, is kept constant. Continuous systems strive to maintain steady state growth conditions, and thus the cell loss due to medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth
20 factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

 It is contemplated that the present invention may be practiced using
25 batch, fed-batch or continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for production of glycerol and glycerol derivatives, such as 1,3-propanediol.

30 In one embodiment, a process for making glycerol, 1,3-propanediol, and/or 3-hydroxypropionic acid from sucrose is provided. The process comprises the steps of culturing a recombinant bacterium, as described

above, in the presence of sucrose, and optionally recovering the glycerol, 1,3-propanediol, and/or 3-hydroxypropionic acid produced. The product may be recovered using methods known in the art. For example, solids may be removed from the fermentation medium by centrifugation, filtration, 5 decantation, or the like. Then, the product may be isolated from the fermentation medium, which has been treated to remove solids as described above, using methods such as distillation, liquid-liquid extraction, or membrane-based separation.

10 EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can 15 ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

GENERAL METHODS

20 Standard recombinant DNA and molecular cloning techniques described in the Examples are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, (1989) (Maniatis) and by T. J. Silhavy, M. L. Bennan, and L. W. 25 Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) and by Ausubel, F. M. *et al.*, *Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987).

The meaning of abbreviations is as follows: "sec" means 30 second(s), "min" means minute(s), "h" means hour(s), "nm" means nanometers, "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "mM" means millimolar, "M" means molar, "g" means gram(s), "μg"

means microgram(s), "bp" means base pair(s), "kbp" means kilobase pair(s), "rpm" means revolutions per minute, "ATCC" means American Type Culture Collection, Manassas, VA.

Media and Culture Conditions:

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following Examples may be found as set out in *Manual of Methods for General Bacteriology* (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials described for the growth and maintenance of bacterial cells may be obtained from Aldrich Chemicals (Milwaukee, WI), BD Diagnostic Systems (Sparks, MD), Life Technologies (Rockville, MD), New England Biolabs (Beverly, MA), or Sigma Chemical Company (St. Louis, MO), unless otherwise specified.

LB (Luria Bertani) medium contains following per liter of medium:
Bacto-tryptone (10 g), Bacto-yeast extract (5 g), and NaCl (10 g).

20 Supplements were added as described in the Examples below. All additions were pre-sterilized before they were added to the medium.

Molecular Biology Techniques:

Restriction enzyme digestions, ligations, transformations, and methods for agarose gel electrophoresis were performed as described in Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press (1989). Polymerase Chain Reactions (PCR) techniques were found in White, B., *PCR Protocols: Current Methods and Applications*, Volume 15 (1993), Humana Press Inc., New York, NY.

30 EXAMPLE 1

Recombinant *E. coli* Containing Putative Sucrose Metabolism Genes

The purpose of this Example was to construct a recombinant *E. coli*

strain containing putative sucrose metabolism genes from *Pseudomonas fluorescens* Pf5 (ATCC® BAA-477), and having the ability to produce 1,3-propanediol (PDO) from sucrose.

P. fluorescens Pf5 genes *lacY* (SEQ ID NO:23) and *scrB* (SEQ ID NO:25), which are adjacent genes, and the promoter region upstream of Pf5 gene *lacY* were obtained by PCR amplification using a high fidelity polymerase, Phusion™ Flash (Finnzymes Oy, Espoo, Finland) using the following conditions: 98 °C/10 sec; 30 cycles of 98 °C/1 sec, 63 °C/30 sec, and 72 °C/30 sec; and then 72 °C/5 min using genomic DNA from *P. fluorescens* Pf5 as template and the primers set forth in SEQ ID NO:56 and SEQ ID NO:57.

An adenosine nucleotide ("A") was added to the 3' ends of the 3063 bp PCR product with Taq polymerase for 10 min at 72 °C. The PCR product was cloned into pBAD-TOPO® (Invitrogen, Carlsbad, CA) following the protocol supplied by the vendor. After transformation of *E. coli* strain TOP10 (Invitrogen) with selection for ampicillin resistance, plasmid DNA was isolated. The presence of the inserted DNA was verified by DNA sequence analysis. One plasmid, named "pBAD-TOPO Pf5 cscB&A 3-5", was transformed into *E. coli* strain TTab pSYCO400/AGRO to generate strain PDO2241. A control plasmid without inserted DNA, named "pBAD-TOPO ctl 3-8", was also transformed into *E. coli* strain TTab pSYCO400/AGRO to generate strain PDO2242.

E. coli strain TTab pSYCO400/AGRO, a PTS minus strain, was constructed as follows. Strain TTab was generated by deletion of the *aldB* gene from strain TT aldA, described in U.S. Patent No. 7,371,558 (Example 17). Briefly, an *aldB* deletion was made by first replacing 1.5 kbp of the coding region of *aldB* in *E. coli* strain MG1655 with the FRT-CmR-FRT cassette of the pKD3 plasmid (Datsenko and Wanner, *Proc. Natl. Acad. Sci. USA* 97:6640-6645, 2000). A replacement cassette was amplified with the primer pair SEQ ID NO:58 and SEQ ID NO:59 using pKD3 as the template. The primer SEQ ID NO:58 contains 80 bp of

homology to the 5'-end of *aldB* and 20 bp of homology to pKD3. Primer SEQ ID NO:59 contains 80 bp of homology to the 3' end of *aldB* and 20 bp homology to pKD3. The PCR products were gel-purified and electroporated into MG1655/pKD46 competent cells (U.S. Patent No. 7,371,558). Recombinant strains were selected on LB plates with 12.5 mg/L of chloramphenicol. The deletion of the *aldB* gene was confirmed by PCR, using the primer pair SEQ ID NO:60 and SEQ ID NO:61. The wild-type strain gave a 1.5 kbp PCR product while the recombinant strain gave a characteristic 1.1 kbp PCR product. A P1 lysate was prepared and used to move the mutation to the TT *aldA* strain to form the TT *aldA*Δ*aldB*::Cm strain. A chloramphenicol-resistant clone was checked by genomic PCR with the primer pair SEQ ID NO:60 and SEQ ID NO:61 to ensure that the mutation was present. The chloramphenicol resistance marker was removed using the FLP recombinase (Datsenko and Wanner, *supra*) to create TTab. Strain TTab was then transformed with pSYCO400/AGRO (set forth in SEQ ID NO:55), described in U.S. Patent No. 7,524,660 (Example 4), to generate strain TTab pSYCO400/AGRO.

As described in the cited references, strain TTab is a derivative of *E. coli* strain FM5 (ATCC® No. 53911) containing the following modifications:

deletion of *glpK*, *gldA*, *ptsHI*, *crp*, *edd*, *arcA*, *mgsA*, *qor*, *ackA*, *pta*, *aldA* and *aldB* genes;

upregulation of *galP*, *glk*, *btuR*, *ppc*, and *yqhD* genes; and

downregulation of *gapA* gene.

Plasmid pSYCO400/AGRO contains genes encoding a glycerol production pathway (*DAR1* and *GPP2*) and genes encoding a glycerol dehydratase and associated reactivating factor (*dhaB123*, *dhaX*, *orfX*, *orfY*), as well as a gene encoding a fructokinase (*scrK*).

30

EXAMPLE 2

Production of 1,3-Propanediol from Sucrose

The purpose of this Example was to demonstrate that an *E. coli*

strain comprising the putative sucrose metabolism genes from *Pseudomonas fluorescens* Pf5 (ATCC® BAA-477) was able to metabolize sucrose and produce 1,3-propanediol (PDO) and glycerol.

E. coli strains PDO2241 and PDO2242, described in Example 1, were grown overnight in L-Broth, Miller's Modification (Teknova, Half Moon Bay, CA) supplemented with 100 mg/L spectinomycin and 100 mg/L ampicillin at 33 °C. These cultures were used to inoculate shake flasks at an optical density of 0.01 units measured at 550 nm in MOPS minimal medium (Teknova, Half Moon Bay, CA) supplemented with 11 g/L sucrose. Vitamin B12 was added to the medium to a concentration of 0.1 mg/L. The cultures were incubated at 34 °C with shaking (225 rpm) for 24 hours, after which time the concentrations of sucrose, glycerol and 1,3-propanediol (PDO) in the broth were determined by high performance liquid chromatography.

Chromatographic separation was achieved using an Aminex HPX-87P column (Bio-Rad, Hercules, CA) with an isocratic mobile phase of distilled-deionized water at a flow rate of 0.5 mL/min and a column temperature of 85 °C. Eluted compounds were quantified by refractive index detection with reference to a standard curve prepared from commercially purchased pure compounds dissolved to known concentrations in MOPS minimal medium. Retention times were sucrose at 12.2 min, 1,3-propanediol at 17.9 min, and glycerol at 23.6 min.

E. coli strain PDO2241 with the putative sucrose metabolism genes from *P. fluorescens* Pf5 metabolized all the sucrose in the medium (0 g/L remaining) and produced 2.55 g/L PDO and 2.83 g/L glycerol. The control strain PDO2242 was unable to metabolize sucrose (10.7 g/L remaining) and produced only negligible PDO (0.06 g/L) and glycerol (0.01 g/L) under these conditions.

30

EXAMPLE 3

Recombinant *E. coli* Containing Putative Sucrose Metabolism Genes

The purpose of this Example was to construct a recombinant *E. coli* strain containing cloned sucrose metabolism genes from *P.*

5 *fluorescens* Pf5.

The plasmid pBAD-TOPO Pf5 cscB&A 3-5, described in Example 1 was transformed into *E. coli* strain FM5 (ATCC® 53911) to generate strain PDO2355. A control plasmid without inserted DNA, pBAD-TOPO
10 ctrl 3-8, was also transformed into *E. coli* strain FM5 to generate strain PDO2350.

E. coli strains PDO2355 and PDO2350 were grown overnight in LB (Luria Bertani) medium containing 100 µg/mL ampicillin at 37 °C. The next day, these cultures were diluted 1:50 in LB medium containing 100 µg/mL ampicillin and grown at 37 °C for 4 hours. These log-phase
15 cultures were diluted 1:100 in the wells of a Bioscreen-C plate (instrument and plates purchased from Growth Curves USA, Piscataway NJ) with MOPS minimal medium (Teknova, Half Moon Bay, CA) containing 10 g/L sucrose. The cultures were grown at 37 °C in triplicate with continuous shaking. Optical density was monitored. At 20 hours
20 after inoculation, the optical density of the PDO2355 culture was 0.584 +/- 0.004, while the optical density of the PDO2350 culture was 0.065 +/- 0.004. These results demonstrate that the control strain was unable to grow with sucrose as sole carbon source, while the strain expressing *P. fluorescens* Pf5 *lacY* and *scrB* was able to grow with sucrose as sole
25 carbon source.

CLAIMS

What is claimed is:

1. A recombinant bacterium comprising in its genome or on at least
5 one recombinant construct:

(a) a nucleotide sequence encoding a polypeptide having sucrose
transporter activity, said polypeptide having at least 95%
sequence identity, based on a Clustal V method of alignment,
when compared to an amino acid sequence as set forth in
10 SEQ ID NO:24; and

(b) a nucleotide sequence encoding a polypeptide having sucrose
hydrolase activity, said polypeptide having at least 95%
sequence identity, based on a Clustal V method of alignment,
when compared to an amino acid sequence as set forth in
15 SEQ ID NO:26;

wherein (a) and (b) are each operably linked to the same or a different
promoter, further wherein said recombinant bacterium is capable of
metabolizing sucrose.

20 2. The recombinant bacterium of claim 1 wherein the nucleotide
sequence encoding the polypeptide having sucrose transporter activity
has at least 95% sequence identity, based on BLASTN method of
alignment, when compared to a nucleotide sequence as set forth in SEQ
ID NO:23.

25 3. The recombinant bacterium of claim 1 wherein the nucleotide
sequence encoding the polypeptide having sucrose hydrolase activity has
at least 95% sequence identity, based on BLASTN method of alignment,
when compared to a nucleotide sequence as set forth in SEQ ID NO:25.

30 4. The recombinant bacterium of claim 1 further comprising in its
genome or on at least one recombinant construct, a nucleotide sequence
encoding a polypeptide having fructokinase activity.

5. The recombinant bacterium of claim 4, wherein the polypeptide having fructokinase activity has at least 95% sequence identity, based on the Clustal V method of alignment, to an amino acid sequence as set forth
5 in SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, or SEQ ID NO:46.

6. The recombinant bacterium of claim 4, wherein the polypeptide having fructokinase activity has an amino acid sequence as set forth in
10 SEQ ID NO:40.

7. The recombinant bacterium of claim 1 wherein said bacterium is selected from the group consisting of the genera: *Escherichia*, *Klebsiella*, *Citrobacter*, and *Aerobacter*.
15

8. The recombinant bacterium of claim 7 wherein said bacterium is *Escherichia coli*.

9. The recombinant bacterium of claim 1 wherein the recombinant
20 bacterium produces 1,3-propanediol, glycerol, and/or 3-hydroxypropionic acid.

10. A process for making glycerol, 1,3-propanediol and/or 3-hydroxypropionic acid from sucrose comprising:
25 a) culturing the recombinant bacterium of claim 9 in the presence of sucrose; and
b) optionally, recovering the glycerol, 1,3-propanediol and/or 3-hydroxypropionic acid produced.

30

CEN5301WOPCTCORRECTEDSEQLIST.txt

SEQUENCE LISTING

<110> JANSSEN BIOTECH, INC.

<120> SERUM MARKERS FOR IDENTIFICATION OF
CUTANEOUS SYSTEMIC SCLEROSIS SUBJECTS

<130> CEN5301WOPCT

<140> TO BE ASSIGNED

<141> 2011-09-27

<150> 61/387580

<151> 2010-09-26

<160> 83

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 277

<212> PRT

<213> Homo sapiens

<400> 1

```

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

```

<210> 2

<211> 1170

<212> PRT

<213> Homo sapiens

<400> 2

```

Met Gly Leu Ala Trp Gly Leu Gly Val Leu Phe Leu Met His Val Cys
 1      5      10      15
Gly Thr Asn Arg Ile Pro Glu Ser Gly Gly Asp Asn Ser Val Phe Asp
 20      25      30
Ile Phe Glu Leu Thr Gly Ala Ala Arg Lys Gly Ser Gly Arg Arg Leu
 35      40      45
Val Lys Gly Pro Asp Pro Ser Ser Pro Ala Phe Arg Ile Glu Asp Ala
 50      55      60
Asn Leu Ile Pro Pro Val Pro Asp Asp Lys Phe Gln Asp Leu Val Asp
 65      70      75
Ala Val Arg Ala Glu Lys Gly Phe Leu Leu Ala Leu Ala Ser Leu Arg Gln
 85      90      95
Met Lys Lys Thr Arg Gly Thr Leu Leu Ala Leu Glu Arg Lys Asp His
100      105      110
Ser Gly Gln Val Phe Ser Val Val Ser Asn Gly Lys Ala Gly Thr Leu
115      120      125
Asp Leu Ser Leu Thr Val Gln Gly Lys Gln His Val Val Ser Val Glu
130      135      140
Glu Ala Leu Leu Ala Thr Gly Gln Trp Lys Ser Ile Thr Leu Phe Val
145      150      155
Gln Glu Asp Arg Ala Gln Leu Tyr Ile Asp Cys Glu Lys Met Glu Asn
165      170      175
Ala Glu Leu Asp Val Pro Ile Gln Ser Val Phe Thr Arg Asp Leu Ala
180      185      190
Ser Ile Ala Arg Leu Arg Ile Ala Lys Gly Gly Val Asn Asp Asn Phe
195      200      205
Gln Gly Val Leu Gln Asn Val Arg Phe Val Phe Gly Thr Thr Pro Glu
210      215      220
Asp Ile Leu Arg Asn Lys Gly Cys Ser Ser Ser Thr Ser Val Leu Leu
225      230      235
Thr Leu Asp Asn Asn Val Val Asn Gly Ser Ser Pro Ala Ile Arg Thr
245      250      255
Asn Tyr Ile Gly His Lys Thr Lys Asp Leu Gln Ala Ile Cys Gly Ile
260      265      270
Ser Cys Asp Glu Leu Ser Ser Met Val Leu Glu Leu Arg Gly Leu Arg
275      280      285
Thr Ile Val Thr Thr Leu Gln Asp Ser Ile Arg Lys Val Thr Glu Glu
290      295      300
Asn Lys Glu Leu Ala Asn Glu Leu Arg Arg Pro Leu Cys Tyr His
305      310      315
Asn Gly Val Gln Tyr Arg Asn Asn Glu Glu Trp Thr Val Asp Ser Cys
325      330      335
Thr Glu Cys His Cys Gln Asn Ser Val Thr Ile Cys Lys Lys Val Ser
340      345      350
Cys Pro Ile Met Pro Cys Ser Asn Ala Thr Val Pro Asp Gly Glu Cys
355      360      365
Cys Pro Arg Cys Trp Pro Ser Asp Ser Ala Asp Asp Gly Trp Ser Pro
370      375      380
Trp Ser Glu Trp Thr Ser Cys Ser Thr Ser Cys Gly Asn Gly Ile Gln
385      390      395
Gln Arg Gly Arg Ser Cys Asp Ser Leu Asn Asn Arg Cys Glu Gly Ser
405      410      415
Ser Val Gln Thr Arg Thr Cys His Ile Gln Glu Cys Asp Lys Arg Phe
420      425      430
Lys Gln Asp Gly Gly Trp Ser His Trp Ser Pro Trp Ser Ser Cys Ser
435      440      445
Val Thr Cys Gly Asp Gly Val Ile Thr Arg Ile Arg Leu Cys Asn Ser
450      455      460
Pro Ser Pro Gln Met Asn Gly Lys Pro Cys Glu Gly Glu Ala Arg Glu
465      470      475
Thr Lys Ala Cys Lys Lys Asp Ala Cys Pro Ile Asn Gly Gly Trp Gly
485      490      495
Pro Trp Ser Pro Trp Asp Ile Cys Ser Val Thr Cys Gly Gly Val
500      505      510
Gln Lys Arg Ser Arg Leu Cys Asn Asn Pro Thr Pro Gln Phe Gly Gly

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

Lys	Asp	515	Val	Gly	Asp	Val	520	Glu	Asn	Gln	Ile	525	Cys	Asn	Lys	Gln
530	Cys					535	Thr					540				
Asp	Cys	Pro	Ile	Asp	Gly	Cys	Leu	Ser	Asn	Pro	Cys	Phe	Ala	Gly	Val	
545				550					555						560	
Lys	Cys	Thr	Ser	Tyr	Pro	Asp	Gly	Ser	Trp	Lys	Cys	Gly	Ala	Cys	Pro	
			565					570						575		
Pro	Gly	Tyr	Ser	Gly	Asn	Gly	Ile	Gln	Cys	Thr	Asp	Val	Asp	Glu	Cys	
		580						585					590			
Lys	Glu	Val	Pro	Asp	Ala	Cys	Phe	Asn	His	Asn	Gly	Glu	His	Arg	Cys	
	595						600					605				
Glu	Asn	Thr	Asp	Pro	Gly	Tyr	Asn	Cys	Leu	Pro	Cys	Pro	Pro	Arg	Phe	
610					615						620					
Thr	Gly	Ser	Gln	Pro	Phe	Gly	Gln	Gly	Val	Glu	His	Ala	Thr	Ala	Asn	
625				630					635						640	
Lys	Gln	Val	Cys	Lys	Pro	Arg	Asn	Pro	Cys	Thr	Asp	Gly	Thr	His	Asp	
			645					650						655		
Cys	Asn	Lys	Asn	Ala	Lys	Cys	Asn	Tyr	Leu	Gly	His	Tyr	Ser	Asp	Pro	
		660						665					670			
Met	Tyr	Arg	Cys	Glu	Cys	Lys	Pro	Gly	Tyr	Ala	Gly	Asn	Gly	Ile	Ile	
		675					680					685				
Cys	Gly	Glu	Asp	Thr	Asp	Leu	Asp	Gly	Trp	Pro	Asn	Glu	Asn	Leu	Val	
	690					695					700					
Cys	Val	Ala	Asn	Ala	Thr	Tyr	His	Cys	Lys	Lys	Asp	Asn	Cys	Pro	Asn	
705				710						715					720	
Leu	Pro	Asn	Ser	Gly	Gln	Glu	Asp	Tyr	Asp	Lys	Asp	Gly	Ile	Gly	Asp	
			725					730					735			
Ala	Cys	Asp	Asp	Asp	Asp	Asp	Asn	Asp	Lys	Ile	Pro	Asp	Asp	Arg	Asp	
		740					745					750				
Asn	Cys	Pro	Phe	His	Tyr	Asn	Pro	Ala	Gln	Tyr	Asp	Tyr	Asp	Arg	Asp	
		755					760					765				
Asp	Val	Gly	Asp	Arg	Cys	Asp	Asn	Cys	Pro	Tyr	Asn	His	Asn	Pro	Asp	
	770					775					780					
Gln	Ala	Asp	Thr	Asp	Asn	Asn	Gly	Glu	Gly	Asp	Ala	Cys	Ala	Ala	Asp	
785				790					795						800	
Ile	Asp	Gly	Asp	Gly	Ile	Leu	Asn	Glu	Arg	Asp	Asn	Cys	Gln	Tyr	Val	
			805					810					815			
Tyr	Asn	Val	Asp	Gln	Arg	Asp	Thr	Asp	Met	Asp	Gly	Val	Gly	Asp	Gln	
		820					825					830				
Cys	Asp	Asn	Cys	Pro	Leu	Glu	His	Asn	Pro	Asp	Gln	Leu	Asp	Ser	Asp	
		835					840					845				
Ser	Asp	Arg	Ile	Gly	Asp	Thr	Cys	Asp	Asn	Asn	Gln	Asp	Ile	Asp	Glu	
	850					855					860					
Asp	Gly	His	Gln	Asn	Asn	Leu	Asp	Asn	Cys	Pro	Tyr	Val	Pro	Asn	Ala	
865				870					875						880	
Asn	Gln	Ala	Asp	His	Asp	Lys	Asp	Gly	Lys	Gly	Asp	Ala	Cys	Asp	His	
			885					890						895		
Asp	Asp	Asp	Asn	Asp	Gly	Ile	Pro	Asp	Lys	Asp	Asn	Cys	Arg	Leu		
		900						905					910			
Val	Pro	Asn	Pro	Asp	Gln	Lys	Asp	Ser	Asp	Gly	Asp	Gly	Arg	Gly	Asp	
		915					920					925				
Ala	Cys	Lys	Asp	Asp	Phe	Asp	His	Asp	Ser	Val	Pro	Asp	Ile	Asp	Asp	
	930				935						940					
Ile	Cys	Pro	Glu	Asn	Val	Asp	Ile	Ser	Glu	Thr	Asp	Phe	Arg	Arg	Phe	
945				950					955						960	
Gln	Met	Ile	Pro	Leu	Asp	Pro	Lys	Gly	Thr	Ser	Gln	Asn	Asp	Pro	Asn	
			965					970						975		
Trp	Val	Val	Arg	His	Gln	Gly	Lys	Glu	Leu	Val	Gln	Thr	Val	Asn	Cys	
		980					985					990				
Asp	Pro	Gly	Leu	Ala	Val	Gly	Tyr	Asp	Glu	Phe	Asn	Ala	Val	Asp	Phe	
		995				1000						1005				
Ser	Gly	Thr	Phe	Phe	Ile	Asn	Thr	Glu	Arg	Asp	Asp	Asp	Tyr	Ala	Gly	
	1010				1015						1020					
Phe	Val	Phe	Gly	Tyr	Gln	Ser	Ser	Ser	Arg	Phe	Tyr	Val	Val	Met	Trp	
1025				1030					1035						1040	
Lys	Gln	Val	Thr	Gln	Ser	Tyr	Trp	Asp	Thr	Asn	Pro	Thr	Arg	Ala	Gln	
			1045					1050						1055		
Gly	Tyr	Ser	Gly	Leu	Ser	Val	Lys	Val	Val	Asn	Ser	Thr	Thr	Gly	Pro	

CEN5301WOPCTCORRECTEDSEQLIST.txt

1060 1065 1070
 Gly Glu His Leu Arg Asn Ala Leu Trp His Thr Gly Asn Thr Pro Gly
 1075 1080 1085
 Gln Val Arg Thr Leu Trp His Asp Pro Arg His Ile Gly Trp Lys Asp
 1090 1095 1100
 Phe Thr Ala Tyr Arg Trp Arg Leu Ser His Arg Pro Lys Thr Gly Phe
 1105 1110 1115
 Ile Arg Val Val Met Tyr Glu Gly Lys Lys Ile Met Ala Asp Ser Gly
 1125 1130 1135
 Pro Ile Tyr Asp Lys Thr Tyr Ala Gly Gly Arg Leu Gly Leu Phe Val
 1140 1145 1150
 Phe Ser Gln Glu Met Val Phe Phe Ser Asp Leu Lys Tyr Glu Cys Arg
 1155 1160 1165
 Asp Pro
 1170

<210> 3
 <211> 1207
 <212> PRT
 <213> Homo sapiens

<400> 3
 Met Leu Leu Thr Leu Ile Ile Leu Leu Pro Val Val Ser Lys Phe Ser
 1 5 10 15
 Phe Val Ser Leu Ser Ala Pro Gln His Trp Ser Cys Pro Glu Gly Thr
 20 25 30
 Leu Ala Gly Asn Gly Asn Ser Thr Cys Val Gly Pro Ala Pro Phe Leu
 35 40 45
 Ile Phe Ser His Gly Asn Ser Ile Phe Arg Ile Asp Thr Glu Gly Thr
 50 55 60
 Asn Tyr Glu Gln Leu Val Val Asp Ala Gly Val Ser Val Ile Met Asp
 65 70 75 80
 Phe His Tyr Asn Glu Lys Arg Ile Tyr Trp Val Asp Leu Glu Arg Gln
 85 90 95
 Leu Leu Gln Arg Val Phe Leu Asn Gly Ser Arg Gln Glu Arg Val Cys
 100 105 110
 Asn Ile Glu Lys Asn Val Ser Gly Met Ala Ile Asn Trp Ile Asn Glu
 115 120 125
 Glu Val Ile Trp Ser Asn Gln Gln Glu Gly Ile Ile Thr Val Thr Asp
 130 135 140
 Met Lys Gly Asn Asn Ser His Ile Leu Leu Ser Ala Leu Lys Tyr Pro
 145 150 155 160
 Ala Asn Val Ala Val Asp Pro Val Glu Arg Phe Ile Phe Trp Ser Ser
 165 170 175
 Glu Val Ala Gly Ser Leu Tyr Arg Ala Asp Leu Asp Gly Val Gly Val
 180 185 190
 Lys Ala Leu Leu Glu Thr Ser Glu Lys Ile Thr Ala Val Ser Leu Asp
 195 200 205
 Val Leu Asp Lys Arg Leu Phe Trp Ile Gln Tyr Asn Arg Glu Gly Ser
 210 215 220
 Asn Ser Leu Ile Cys Ser Cys Asp Tyr Asp Gly Gly Ser Val His Ile
 225 230 235 240
 Ser Lys His Pro Thr Gln His Asn Leu Phe Ala Met Ser Leu Phe Gly
 245 250 255
 Asp Arg Ile Phe Tyr Ser Thr Trp Lys Met Lys Thr Ile Trp Ile Ala
 260 265 270
 Asn Lys His Thr Gly Lys Asp Met Val Arg Ile Asn Leu His Ser Ser
 275 280 285
 Phe Val Pro Leu Gly Glu Leu Lys Val Val His Pro Leu Ala Gln Pro
 290 295 300
 Lys Ala Glu Asp Asp Thr Trp Glu Pro Glu Gln Lys Leu Cys Lys Leu
 305 310 315 320
 Arg Lys Gly Asn Cys Ser Ser Thr Val Cys Gly Gln Asp Leu Gln Ser
 325 330 335
 His Leu Cys Met Cys Ala Glu Gly Tyr Ala Leu Ser Arg Asp Arg Lys
 340 345 350
 Tyr Cys Glu Asp Val Asn Glu Cys Ala Phe Trp Asn His Gly Cys Thr

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

Leu Gly 355 Lys Asn Thr Pro 360 Gly Ser Tyr Tyr Cys 365 Thr Cys Pro Val
370 Cys Lys Val Leu Leu Pro 375 Asp Gly Lys Arg Cys 380 His Gln Leu Val Ser
385 Phe Val Leu Leu Pro 390 Asp Gly Lys Arg Cys 395 His Gln Leu Val Ser
Cys Pro Arg Asn Val Ser Glu Cys Ser His 410 Asp Cys Val Leu Thr Ser
405 Phe Cys Pro Glu 425 Gly Ser Val Leu Glu 415 Arg Asp
Glu Gly Pro Leu 420 Cys Phe Cys Pro Glu 425 Gly Ser Val Leu Glu 430 Arg Asp
Gly Lys Thr Cys Ser Gly Cys Ser Ser Pro Asp Asn Gly Gly Cys Ser
435 Leu Cys Val Pro Leu Ser Pro Val Ser Trp Glu Cys Asp Cys Phe
450 Gly Tyr Asp Leu Gln 455 Leu Asp Glu Lys Ser Cys Ala Ala Ser Gly
465 Pro Gln Pro Phe Leu Leu Phe Ala Asn Ser Gln Asp Ile Arg His Met
485 Thr Asp Tyr Gly Thr 505 Leu Leu Ser Gln Gln Met Gly
His Phe Asp Gly 500 Thr Asp Tyr Gly Thr 505 Leu Leu Ser Gln Gln Met Gly
Met Val Tyr 515 Ala Leu Asp His Asp 520 Pro Val Glu Asn Lys 525 Ile Tyr Phe
Ala His Thr Ala Leu Lys Trp Ile Glu Arg Ala Asn Met Asp Gly Ser
530 Gln Arg Glu Arg Leu Ile Glu Glu Gly Val Asp Val Pro Glu Gly Leu
545 Ala Val Asp Trp Ile 550 Gly Arg Arg Phe Tyr 555 Trp Thr Asp Arg Gly Lys
565 Ser Leu Ile Gly Arg Ser Asp Leu Asn Gly Lys Arg Ser Lys Ile Ile
580 Thr Lys Glu Asn Ile Ser Gln Pro Arg Gly Ile Ala Val His Pro Met
595 Ala Lys Arg Leu Phe Trp Thr 600 Thr Gly Ile Asn Pro Arg Ile Glu
610 Ser Ser Ser Leu Gln Gly Leu Gly Arg Leu Val Ile Ala Ser Ser Asp
625 Leu Ile Trp Pro Ser Gly Ile Thr Ile Asp Phe Leu Thr Asp Lys Leu
645 Tyr Trp Cys Asp Ala Lys Gln Ser Val Ile Glu Met Ala Asn Leu Asp
660 Gly Ser Lys Arg Arg Arg Leu Thr Gln Asn Asp Val Gly His Pro Phe
675 Ala Val Ala Val Phe Glu Asp Tyr Val Trp Phe Ser Asp Trp Ala Met
690 Pro Ser Val Met Arg Val Asn Lys Arg Thr Gly Lys Asp Arg Val Arg
705 Leu Gln Gly Ser Met Leu Lys Pro Ser Ser Leu Val Val Val His Pro
725 Leu Ala Lys Pro Gly Ala Asp Pro Cys 730 Leu Tyr Gln Asn Gly Gly Cys
740 Glu His Ile Cys Lys Lys Arg Leu Gly Thr Ala Trp Cys 750 Ser Cys Arg
755 Glu Gly Phe Met Lys Ala Ser Asp Gly Lys Thr Cys Leu Ala Leu Asp
770 Gly His Gln Leu Leu Ala Gly Gly Glu Val Asp Leu Lys Asn Gln Val
785 Thr Pro Leu Asp Ile Leu Ser Lys Thr Arg Val Ser Glu Asp Asn Ile
805 Thr Glu Ser Gln His Met Leu Val Ala Glu Ile Met Val Ser Asp Gln
820 Asp Asp Cys Ala Pro Val Gly Cys Ser Met Tyr Ala Arg Cys Ile Ser
835 Glu Gly Glu Asp Ala Thr Cys Gln Cys Leu Lys Gly Phe Ala Gly Asp
850 Gly Lys Leu Cys Ser Asp Ile Asp Glu Cys Glu Met Gly Val Pro Val
865 Cys Pro Pro Ala Ser 870 Lys Cys Ile Asn Thr Glu Gly Gly Tyr Val
885 Cys Arg Cys Ser Glu Gly Tyr Gln Gly Asp Gly Ile His Cys Leu Asp
890

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

          900          905          910
Ile Asp Glu Cys Gln Leu Gly Glu His Ser Cys Gly Glu Asn Ala Ser
      915          920          925
Cys Thr Asn Thr Glu Gly Gly Tyr Thr Cys Met Cys Ala Gly Arg Leu
      930          935          940
Ser Glu Pro Gly Leu Ile Cys Pro Asp Ser Thr Pro Pro Pro His Leu
945      950      955
Arg Glu Asp Asp His Tyr Ser Val Arg Asn Ser Asp Ser Glu Cys
      965      970      975
Pro Leu Ser His Asp Gly Tyr Cys Leu His Asp Gly Val Cys Met Tyr
      980      985      990
Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn Cys Val Val Gly Tyr Ile
      995      1000      1005
Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys Trp Trp Glu Leu Arg His
      1010      1015      1020
Ala Gly His Gly Gln Gln Gln Lys Val Ile Val Val Ala Val Cys Val
1025      1030      1035      1040
Val Val Leu Val Met Leu Leu Leu Leu Ser Leu Trp Gly Ala His Tyr
      1045      1050      1055
Tyr Arg Thr Gln Lys Leu Leu Ser Lys Asn Pro Lys Asn Pro Tyr Glu
      1060      1065      1070
Glu Ser Ser Arg Asp Val Arg Ser Arg Arg Pro Ala Asp Thr Glu Asp
      1075      1080      1085
Gly Met Ser Ser Cys Pro Gln Pro Trp Phe Val Val Ile Lys Glu His
      1090      1095      1100
Gln Asp Leu Lys Asn Gly Gly Gln Pro Val Ala Gly Glu Asp Gly Gln
1105      1110      1115      1120
Ala Ala Asp Gly Ser Met Gln Pro Thr Ser Trp Arg Gln Glu Pro Gln
      1125      1130      1135
Leu Cys Gly Met Gly Thr Glu Gln Gly Cys Trp Ile Pro Val Ser Ser
      1140      1145      1150
Asp Lys Gly Ser Cys Pro Gln Val Met Glu Arg Ser Phe His Met Pro
      1155      1160      1165
Ser Tyr Gly Thr Gln Thr Leu Glu Gly Gly Val Glu Lys Pro His Ser
      1170      1175      1180
Leu Leu Ser Ala Asn Pro Leu Trp Gln Gln Arg Ala Leu Asp Pro Pro
1185      1190      1195      1200
His Gln Met Glu Leu Thr Gln
      1205

```

<210> 4
 <211> 116
 <212> PRT
 <213> Homo sapiens

```

<400> 4
Met Asp Tyr Tyr Arg Lys Tyr Ala Ala Ile Phe Leu Val Thr Leu Ser
 1      5      10      15
Val Phe Leu His Val Leu His Ser Ala Pro Asp Val Gln Asp Cys Pro
      20      25      30
Glu Cys Thr Leu Gln Glu Asn Pro Phe Phe Ser Gln Pro Gly Ala Pro
      35      40      45
Ile Leu Gln Cys Met Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro
 50      55      60
Leu Arg Ser Lys Lys Thr Met Leu Val Gln Lys Asn Val Thr Ser Glu
65      70      75      80
Ser Thr Cys Cys Val Ala Lys Ser Tyr Asn Arg Val Thr Val Met Gly
      85      90      95
Gly Phe Lys Val Glu Asn His Thr Ala Cys His Cys Ser Thr Cys Tyr
      100      105      110
Tyr His Lys Ser
      115

```

<210> 5
 <211> 241
 <212> PRT

<213> Homo sapiens

<400> 5

```

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

```

<210> 6

<211> 91

<212> PRT

<213> Homo sapiens

<400> 6

```

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

```

<210> 7

<211> 247

<212> PRT

<213> Homo sapiens

<400> 7

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 8
 <211> 160
 <212> PRT
 <213> Homo sapiens

```

<400> 8
Met  Val  Pro  Ser  Ala  Gly  Gln  Leu  Ala  Leu  Phe  Ala  Leu  Gly  Ile  Val
   1      5      10      15
Leu  Ala  Ala  Cys  Gln  Ala  Leu  Glu  Asn  Ser  Thr  Ser  Pro  Leu  Ser  Ala
      20      25      30
Asp  Pro  Pro  Val  Ala  Ala  Ala  Val  Val  Ser  His  Phe  Asn  Asp  Cys  Pro
      35      40      45
Asp  Ser  His  Thr  Gln  Phe  Cys  Phe  His  Gly  Thr  Cys  Arg  Phe  Leu  Val
      50      55      60
Gln  Glu  Asp  Lys  Pro  Ala  Cys  Val  Cys  His  Ser  Gly  Tyr  Val  Gly  Ala
   65      70      75      80
Arg  Cys  Glu  His  Ala  Asp  Leu  Leu  Ala  Val  Val  Ala  Ala  Ser  Gln  Lys
      85      90      95
Lys  Gln  Ala  Ile  Thr  Ala  Leu  Val  Val  Ser  Ile  Val  Ala  Leu  Ala
      100      105      110
Val  Leu  Ile  Ile  Thr  Cys  Val  Leu  Ile  His  Cys  Cys  Gln  Val  Arg  Lys
      115      120      125
His  Cys  Glu  Trp  Cys  Arg  Ala  Leu  Ile  Cys  Arg  His  Glu  Lys  Pro  Ser
      130      135      140
Ala  Leu  Leu  Lys  Gly  Arg  Thr  Ala  Cys  Cys  His  Ser  Glu  Thr  Val  Val
   145      150      155      160

```

<210> 9
 <211> 169
 <212> PRT
 <213> Homo sapiens

```

<400> 9
Met  Thr  Ala  Gly  Arg  Arg  Met  Glu  Met  Leu  Cys  Ala  Gly  Arg  Val  Pro
   1      5      10      15
Ala  Leu  Leu  Leu  Cys  Leu  Gly  Phe  His  Leu  Leu  Gln  Ala  Val  Leu  Ser
      20      25      30
Thr  Thr  Val  Ile  Pro  Ser  Cys  Ile  Pro  Gly  Glu  Ser  Ser  Asp  Asn  Cys

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 10
 <211> 208
 <212> PRT
 <213> Homo sapiens

```

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

```

<210> 11
 <211> 252
 <212> PRT
 <213> Homo sapiens

```

<400> 11
Met Arg Ala Pro Leu Leu Pro Pro Ala Pro Val Val Leu Ser Leu Leu
  1      5      10      15
Ile Leu Gly Ser Gly His Tyr Ala Ala Gly Leu Asp Leu Asn Asp Thr
      20      25      30
Tyr Ser Gly Lys Arg Glu Pro Phe Ser Gly Asp His Ser Ala Asp Gly
      35      40      45
Phe Glu Val Thr Ser Arg Ser Glu Met Ser Ser Gly Ser Glu Ile Ser
      50      55      60
Pro Val Ser Glu Met Pro Ser Ser Ser Glu Pro Ser Ser Gly Ala Asp

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 12
 <211> 224
 <212> PRT
 <213> Homo sapiens

```

<400> 12
Met Glu Lys Leu Leu Cys Phe Leu Val Leu Thr Ser Leu Ser His Ala
 1      5      10      15
Phe Gly Gln Thr Asp Met Ser Arg Lys Ala Phe Val Phe Pro Lys Glu
      20      25      30
Ser Asp Thr Ser Tyr Val Ser Leu Lys Ala Pro Leu Thr Lys Pro Leu
      35      40      45
Lys Ala Phe Thr Val Cys Leu His Phe Tyr Thr Glu Leu Ser Ser Thr
      50      55      60
Arg Gly Tyr Ser Ile Phe Ser Tyr Ala Thr Lys Arg Gln Asp Asn Glu
65      70      75      80
Ile Leu Ile Phe Trp Ser Lys Asp Ile Gly Tyr Ser Phe Thr Val Gly
      85      90      95
Gly Ser Glu Ile Leu Phe Glu Val Pro Glu Val Thr Val Ala Pro Val
      100      105      110
His Ile Cys Thr Ser Trp Glu Ser Ala Ser Gly Ile Val Glu Phe Trp
      115      120      125
Val Asp Gly Lys Pro Arg Val Arg Lys Ser Leu Lys Lys Gly Tyr Thr
      130      135      140
Val Gly Ala Glu Ala Ser Ile Ile Leu Gly Gln Glu Gln Asp Ser Phe
145      150      155      160
Gly Gly Asn Phe Glu Gly Ser Gln Ser Leu Val Gly Asp Ile Gly Asn
      165      170      175
Val Asn Met Trp Asp Phe Val Leu Ser Pro Asp Glu Ile Asn Thr Ile
      180      185      190
Tyr Leu Gly Gly Pro Phe Ser Pro Asn Val Leu Asn Trp Arg Ala Leu
      195      200      205
Lys Tyr Glu Val Gln Gly Glu Val Phe Thr Lys Pro Gln Leu Trp Pro
210      215      220

```

<210> 13
 <211> 98
 <212> PRT
 <213> Homo sapiens

```

<400> 13
Met Lys Val Ser Ala Val Leu Leu Cys Leu Leu Leu Met Thr Ala Ala

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 14
 <211> 150
 <212> PRT
 <213> Homo sapiens

```

<400> 14
Met Asn Leu Trp Leu Leu Ala Cys Leu Val Ala Gly Phe Leu Gly Ala
1           5           10           15
Trp Ala Pro Ala Val His Thr Gln Gly Val Phe Glu Asp Cys Cys Leu
20          25          30
Ala Tyr His Tyr Pro Ile Gly Trp Ala Val Leu Arg Arg Ala Trp Thr
35          40          45
Tyr Arg Ile Gln Glu Val Ser Gly Ser Cys Asn Leu Pro Ala Ala Ile
50          55          60
Phe Tyr Leu Pro Lys Arg His Arg Lys Val Cys Gly Asn Pro Lys Ser
65          70          75          80
Arg Glu Val Gln Arg Ala Met Lys Leu Leu Asp Ala Arg Asn Lys Val
85          90          95
Phe Ala Lys Leu His His Asn Thr Gln Thr Phe Gln Ala Gly Pro His
100         105         110
Ala Val Lys Lys Leu Ser Ser Gly Asn Ser Lys Leu Ser Ser Ser Lys
115         120         125
Phe Ser Asn Pro Ile Ser Ser Ser Lys Arg Asn Val Ser Leu Leu Ile
130         135         140
Ser Ala Asn Ser Gly Leu
145         150

```

<210> 15
 <211> 114
 <212> PRT
 <213> Homo sapiens

```

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

```

<210> 16

<211> 134

<212> PRT

<213> Homo sapiens

<400> 16

```

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

```

<210> 17

<211> 155

<212> PRT

<213> Homo sapiens

<400> 17

```

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

```

<210> 18

<211> 170

<212> PRT

<213> Homo sapiens

<400> 18

```

Met Pro Val Met Arg Leu Phe Pro Cys Phe Leu Gln Leu Leu Ala Gly
 1      5      10      15
Leu Ala Leu Pro Ala Val Pro Pro Gln Gln Trp Ala Leu Ser Ala Gly
 20      25      30
Asn Gly Ser Ser Glu Val Glu Val Val Pro Phe Gln Glu Val Trp Gly
 35      40      45
Arg Ser Tyr Cys Arg Ala Leu Glu Arg Leu Val Asp Val Val Ser Glu
 50      55      60
Tyr Pro Ser Glu Val Glu His Met Phe Ser Pro Ser Cys Val Ser Leu

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

65      70      75      80
Leu Arg Cys Thr Gly Cys Cys Gly Asp Glu Asn Leu His Cys Val Pro
      85      90      95
Val Glu Thr Ala Asn Val Thr Met Gln Leu Leu Lys Ile Arg Ser Gly
      100      105      110
Asp Arg Pro Ser Tyr Val Glu Leu Thr Phe Ser Gln His Val Arg Cys
      115      120      125
Glu Cys Arg Pro Leu Arg Glu Lys Met Lys Pro Glu Arg Arg Arg Pro
      130      135      140
Lys Gly Arg Gly Lys Arg Arg Arg Glu Lys Gln Arg Pro Thr Asp Cys
      145      150      155      160
His Leu Cys Gly Asp Ala Val Pro Arg Arg
      165      170

```

<210> 19
 <211> 469
 <212> PRT
 <213> Homo sapiens

```

<400> 19
Met His Ser Phe Pro Pro Leu Leu Leu Leu Leu Phe Trp Gly Val Val
  1      5      10      15
Ser His Ser Phe Pro Ala Thr Leu Glu Thr Gln Glu Gln Asp Val Asp
      20      25      30
Leu Val Gln Lys Tyr Leu Glu Lys Tyr Tyr Asn Leu Lys Asn Asp Gly
      35      40      45
Arg Gln Val Glu Lys Arg Arg Asn Ser Gly Pro Val Val Glu Lys Leu
      50      55      60
Lys Gln Met Gln Glu Phe Phe Gly Leu Lys Val Thr Gly Lys Pro Asp
      65      70      75      80
Ala Glu Thr Leu Lys Val Met Lys Gln Pro Arg Cys Gly Val Pro Asp
      85      90      95
Val Ala Gln Phe Val Leu Thr Glu Gly Asn Pro Arg Trp Glu Gln Thr
      100      105      110
His Leu Thr Tyr Arg Ile Glu Asn Tyr Thr Pro Asp Leu Pro Arg Ala
      115      120      125
Asp Val Asp His Ala Ile Glu Lys Ala Phe Gln Leu Trp Ser Asn Val
      130      135      140
Thr Pro Leu Thr Phe Thr Lys Val Ser Glu Gly Gln Ala Asp Ile Met
      145      150      155      160
Ile Ser Phe Val Arg Gly Asp His Arg Asp Asn Ser Pro Phe Asp Gly
      165      170      175
Pro Gly Gly Asn Leu Ala His Ala Phe Gln Pro Gly Pro Gly Ile Gly
      180      185      190
Gly Asp Ala His Phe Asp Glu Asp Glu Arg Trp Thr Asn Asn Phe Arg
      195      200      205
Glu Tyr Asn Leu His Arg Val Ala Ala His Glu Leu Gly His Ser Leu
      210      215      220
Gly Leu Ser His Ser Thr Asp Ile Gly Ala Leu Met Tyr Pro Ser Tyr
      225      230      235      240
Thr Phe Ser Gly Asp Val Gln Leu Ala Gln Asp Asp Ile Asp Gly Ile
      245      250      255
Gln Ala Ile Tyr Gly Arg Ser Gln Asn Pro Val Gln Pro Ile Gly Pro
      260      265      270
Gln Thr Pro Lys Ala Cys Asp Ser Lys Leu Thr Phe Asp Ala Ile Thr
      275      280      285
Thr Ile Arg Gly Glu Val Met Phe Phe Lys Asp Arg Phe Tyr Met Arg
      290      295      300
Thr Asn Pro Phe Tyr Pro Glu Val Glu Leu Asn Phe Ile Ser Val Phe
      305      310      315      320
Trp Pro Gln Leu Pro Asn Gly Leu Glu Ala Ala Tyr Glu Phe Ala Asp
      325      330      335
Arg Asp Glu Val Arg Phe Phe Lys Gly Asn Lys Tyr Trp Ala Val Gln
      340      345      350
Gly Gln Asn Val Leu His Gly Tyr Pro Lys Asp Ile Tyr Ser Ser Phe
      355      360      365
Gly Phe Pro Arg Thr Val Lys His Ile Asp Ala Ala Leu Ser Glu Glu

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

370
Asn Thr Gly Lys Thr Tyr 375 Phe Phe Val Ala Asn 380 Lys Tyr Trp Arg Tyr
385
Asp Glu Tyr Lys Arg Ser Met Asp Pro Gly Tyr Pro Lys Met Ile Ala
405
His Asp Phe Pro Gly Ile Gly His Lys Val Asp Ala Val Phe Met Lys
420
Asp Gly Phe Phe Tyr Phe Phe His 425 Gly Thr Arg Gln Tyr Lys Phe Asp
435
Pro Lys Thr Lys Arg Ile Leu Thr Leu Gln Lys Ala Asn Ser Trp Phe
450
Asn Cys Arg Lys Asn
465

```

<210> 20
 <211> 402
 <212> PRT
 <213> Homo sapiens

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

370 375 380
 Val Arg His Asn Pro Thr Gly Thr Val Leu Phe Met Gly Gln Val Met
 385 390 395 400
 Glu Pro

<210> 21
 <211> 120
 <212> PRT
 <213> Homo sapiens

<400> 21
 Met His Pro Leu Leu Asn Pro Leu Leu Leu Ala Leu Gly Leu Met Ala
 1 5 10 15
 Leu Leu Leu Thr Thr Val Ile Ala Leu Thr Cys Leu Gly Gly Phe Ala
 20 25 30
 Ser Pro Gly Pro Val Pro Pro Ser Thr Ala Leu Arg Glu Leu Ile Glu
 35 40 45
 Glu Leu Val Asn Ile Thr Gln Asn Gln Lys Ala Pro Leu Cys Asn Gly
 50 55 60
 Ser Met Val Trp Ser Ile Asn Leu Thr Ala Gly Met Tyr Cys Ala Ala
 65 70 75 80
 Leu Glu Ser Leu Ile Asn Val Ser Gly Cys Ser Ala Ile Glu Lys Thr
 85 90 95
 Gln Arg Met Leu Ser Gly Phe Cys Pro His Lys Val Ser Ala Gly Gln
 100 105 110
 Phe Ser Ser Leu His Val Arg Asp
 115 120

<210> 22
 <211> 152
 <212> PRT
 <213> Homo sapiens

<400> 22
 Met Ser Arg Leu Pro Val Leu Leu Leu Leu Gln Leu Leu Val Arg Pro
 1 5 10 15
 Gly Leu Gln Ala Pro Met Thr Gln Thr Thr Pro Leu Lys Thr Ser Trp
 20 25 30
 Val Asn Cys Ser Asn Met Ile Asp Glu Ile Ile Thr His Leu Lys Gln
 35 40 45
 Pro Pro Leu Pro Leu Leu Asp Phe Asn Asn Leu Asn Gly Glu Asp Gln
 50 55 60
 Asp Ile Leu Met Glu Asn Asn Leu Arg Arg Pro Asn Leu Glu Ala Phe
 65 70 75 80
 Asn Arg Ala Val Lys Ser Leu Gln Asn Ala Ser Ala Ile Glu Ser Ile
 85 90 95
 Leu Lys Asn Leu Leu Pro Cys Leu Pro Leu Ala Thr Ala Ala Pro Thr
 100 105 110
 Arg His Pro Ile His Ile Lys Asp Gly Asp Trp Asn Glu Phe Arg Arg
 115 120 125
 Lys Leu Thr Phe Tyr Leu Lys Thr Leu Glu Asn Ala Gln Ala Gln Gln
 130 135 140
 Thr Thr Leu Ser Leu Ala Ile Phe
 145 150

<210> 23
 <211> 99
 <212> PRT
 <213> Homo sapiens

<400> 23
 Met Lys Val Ser Ala Ala Leu Leu Cys Leu Leu Leu Met Ala Ala Thr
 1 5 10 15
 Phe Ser Pro Gln Gly Leu Ala Gln Pro Asp Ser Val Ser Ile Pro Ile
 Page 15

CEN5301WOPCTCORRECTEDSEQLIST.txt

20 25 30
 Thr Cys Cys Phe Asn Val Ile Asn Arg Lys Ile Pro Ile Gln Arg Leu
 35 40 45
 Glu Ser Tyr Thr Arg Ile Thr Asn Ile Gln Cys Pro Lys Glu Ala Val
 50 55 60
 Ile Phe Lys Thr Lys Arg Gly Lys Glu Val Cys Ala Asp Pro Lys Glu
 65 70 75 80
 Arg Trp Val Arg Asp Ser Met Lys His Leu Asp Gln Ile Phe Gln Asn
 85 90 95
 Leu Lys Pro

<210> 24
 <211> 317
 <212> PRT
 <213> Homo sapiens

<400> 24
 Met Lys Val Leu Trp Ala Ala Leu Leu Val Thr Phe Leu Ala Gly Cys
 1 5 10 15
 Gln Ala Lys Val Glu Gln Ala Val Glu Thr Glu Pro Glu Pro Glu Leu
 20 25 30
 Arg Gln Gln Thr Glu Trp Gln Ser Gly Gln Arg Trp Glu Leu Ala Leu
 35 40 45
 Gly Arg Phe Trp Asp Tyr Leu Arg Trp Val Gln Thr Leu Ser Glu Gln
 50 55 60
 Val Gln Glu Glu Leu Leu Ser Ser Gln Val Thr Gln Glu Leu Arg Ala
 65 70 75 80
 Leu Met Asp Glu Thr Met Lys Glu Leu Lys Ala Tyr Lys Ser Glu Leu
 85 90 95
 Glu Glu Gln Leu Thr Pro Val Ala Glu Glu Thr Arg Ala Arg Leu Ser
 100 105 110
 Lys Glu Leu Gln Ala Ala Gln Ala Arg Leu Gly Ala Asp Met Glu Asp
 115 120 125
 Val Cys Gly Arg Leu Val Gln Tyr Arg Gly Glu Val Gln Ala Met Leu
 130 135 140
 Gly Gln Ser Thr Glu Glu Leu Arg Val Arg Leu Ala Ser His Leu Arg
 145 150 155 160
 Lys Leu Arg Lys Arg Leu Leu Arg Asp Ala Asp Asp Leu Gln Lys Arg
 165 170 175
 Leu Ala Val Tyr Gln Ala Gly Ala Arg Glu Gly Ala Glu Arg Gly Leu
 180 185 190
 Ser Ala Ile Arg Glu Arg Leu Gly Pro Leu Val Glu Gln Gly Arg Val
 195 200 205
 Arg Ala Ala Thr Val Gly Ser Leu Ala Gly Gln Pro Leu Gln Glu Arg
 210 215 220
 Ala Gln Ala Trp Gly Glu Arg Leu Arg Ala Arg Met Glu Glu Met Gly
 225 230 235 240
 Ser Arg Thr Arg Asp Arg Leu Asp Glu Val Lys Glu Gln Val Ala Glu
 245 250 255
 Val Arg Ala Lys Leu Glu Glu Gln Ala Gln Gln Ile Arg Leu Gln Ala
 260 265 270
 Glu Ala Phe Gln Ala Arg Leu Lys Ser Trp Phe Glu Pro Leu Val Glu
 275 280 285
 Asp Met Gln Arg Gln Trp Ala Gly Leu Val Glu Lys Val Gln Ala Ala
 290 295 300
 Val Gly Thr Ser Ala Ala Pro Val Pro Ser Asp Asn His
 305 310 315

<210> 25
 <211> 97
 <212> PRT
 <213> Homo sapiens

<400> 25
 Met Lys Val Ser Ala Ala Leu Leu Trp Leu Leu Leu Ile Ala Ala Ala
 Page 16

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 26
 <211> 99
 <212> PRT
 <213> Homo sapiens

```

<400> 26
Met Lys Val Ser Ala Ala Leu Leu Cys Leu Leu Leu Ile Ala Ala Thr
1           5           10           15
Phe Ile Pro Gln Gly Leu Ala Gln Pro Asp Ala Ile Asn Ala Pro Val
20           25           30
Thr Cys Cys Tyr Asn Phe Thr Asn Arg Lys Ile Ser Val Gln Arg Leu
35           40           45
Ala Ser Tyr Arg Arg Ile Thr Ser Ser Lys Cys Pro Lys Glu Ala Val
50           55           60
Ile Phe Lys Thr Ile Val Ala Lys Glu Ile Cys Ala Asp Pro Lys Gln
65           70           75           80
Lys Trp Val Gln Asp Ser Met Asp His Leu Asp Lys Gln Thr Gln Thr
85           90           95
Pro Lys Thr

```

<210> 27
 <211> 314
 <212> PRT
 <213> Homo sapiens

```

<400> 27
Met Arg Ile Ala Val Ile Cys Phe Cys Leu Leu Gly Ile Thr Cys Ala
1           5           10           15
Ile Pro Val Lys Gln Ala Asp Ser Gly Ser Ser Glu Glu Lys Gln Leu
20           25           30
Tyr Asn Lys Tyr Pro Asp Ala Val Ala Thr Trp Leu Asn Pro Asp Pro
35           40           45
Ser Gln Lys Gln Asn Leu Leu Ala Pro Gln Asn Ala Val Ser Ser Glu
50           55           60
Glu Thr Asn Asp Phe Lys Gln Glu Thr Leu Pro Ser Lys Ser Asn Glu
65           70           75           80
Ser His Asp His Met Asp Asp Met Asp Asp Glu Asp Asp Asp Asp His
85           90           95
Val Asp Ser Gln Asp Ser Ile Asp Ser Asn Asp Ser Asp Asp Val Asp
100           105           110
Asp Thr Asp Asp Ser His Gln Ser Asp Glu Ser His His Ser Asp Glu
115           120           125
Ser Asp Glu Leu Val Thr Asp Phe Pro Thr Asp Leu Pro Ala Thr Glu
130           135           140
Val Phe Thr Pro Val Val Pro Thr Val Asp Thr Tyr Asp Gly Arg Gly
145           150           155           160
Asp Ser Val Val Tyr Gly Leu Arg Ser Lys Ser Lys Lys Phe Arg Arg
165           170           175
Pro Asp Ile Gln Tyr Pro Asp Ala Thr Asp Glu Asp Ile Thr Ser His
180           185           190
Met Glu Ser Glu Glu Leu Asn Gly Ala Tyr Lys Ala Ile Pro Val Ala

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

195 200 205
 Gln Asp Leu Asn Ala Pro Ser Asp Trp Asp Ser Arg Gly Lys Asp Ser
 210 215 220
 Tyr Glu Thr Ser Gln Leu Asp Asp Gln Ser Ala Glu Thr His Ser His
 225 230 235 240
 Lys Gln Ser Arg Leu Tyr Lys Arg Lys Ala Asn Asp Glu Ser Asn Glu
 245 250 255
 His Ser Asp Val Ile Asp Ser Gln Glu Leu Ser Lys Val Ser Arg Glu
 260 265 270
 Phe His Ser His Glu Phe His Ser His Glu Asp Met Leu Val Val Asp
 275 280 285
 Pro Lys Ser Lys Glu Glu Asp Lys His Leu Lys Phe Arg Ile Ser His
 290 295 300
 Glu Leu Asp Ser Ala Ser Glu Val Asn
 305 310

<210> 28

<211> 583

<212> PRT

<213> Homo sapiens

<400> 28

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

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 29
 <211> 92
 <212> PRT
 <213> Homo sapiens

```

<400> 29
Met Thr Lys Leu Glu Glu His Leu Glu Gly Ile Val Asn Ile Phe His
1      5      10      15
Gln Tyr Ser Val Arg Lys Gly His Phe Asp Thr Leu Ser Lys Gly Glu
20      25      30
Leu Lys Gln Leu Leu Thr Lys Glu Leu Ala Asn Thr Ile Lys Asn Ile
35      40      45
Lys Asp Lys Ala Val Ile Asp Glu Ile Phe Gln Gly Leu Asp Ala Asn
50      55      60
Gln Asp Glu Gln Val Asp Phe Gln Glu Phe Ile Ser Leu Val Ala Ile
65      70      75      80
Ala Leu Lys Ala Ala His Tyr His Thr His Lys Glu
85      90

```

<210> 30
 <211> 92
 <212> PRT
 <213> Homo sapiens

```

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

```

<210> 31
 <211> 89
 <212> PRT
 <213> Homo sapiens

<400> 31
 Met Lys Gly Leu Ala Ala Ala Leu Leu Val Leu Val Cys Thr Met Ala
 1 5 10 15
 Leu Cys Ser Cys Ala Gln Val Gly Thr Asn Lys Glu Leu Cys Cys Leu
 20 25 30
 Val Tyr Thr Ser Trp Gln Ile Pro Gln Lys Phe Ile Val Asp Tyr Ser
 35 40 45
 Glu Thr Ser Pro Gln Cys Pro Lys Pro Gly Val Ile Leu Leu Thr Lys
 50 55 60
 Arg Gly Arg Gln Ile Cys Ala Asp Pro Asn Lys Lys Trp Val Gln Lys
 65 70 75 80
 Tyr Ile Ser Asp Leu Lys Leu Asn Ala
 85

<210> 32
 <211> 728
 <212> PRT
 <213> Homo sapiens

<400> 32
 Met Trp Val Thr Lys Leu Leu Pro Ala Leu Leu Leu Gln His Val Leu
 1 5 10 15
 Leu His Leu Leu Leu Leu Pro Ile Ala Ile Pro Tyr Ala Glu Gly Gln
 20 25 30
 Arg Lys Arg Arg Asn Thr Ile His Glu Phe Lys Lys Ser Ala Lys Thr
 35 40 45
 Thr Leu Ile Lys Ile Asp Pro Ala Leu Lys Ile Lys Thr Lys Lys Val
 50 55 60
 Asn Thr Ala Asp Gln Cys Ala Asn Arg Cys Thr Arg Asn Lys Gly Leu
 65 70 75 80
 Pro Phe Thr Cys Lys Ala Phe Val Phe Asp Lys Ala Arg Lys Gln Cys
 85 90 95
 Leu Trp Phe Pro Phe Asn Ser Met Ser Ser Gly Val Lys Lys Glu Phe
 100 105 110
 Gly His Glu Phe Asp Leu Tyr Glu Asn Lys Asp Tyr Ile Arg Asn Cys
 115 120 125
 Ile Ile Gly Lys Gly Arg Ser Tyr Lys Gly Thr Val Ser Ile Thr Lys
 130 135 140
 Ser Gly Ile Lys Cys Gln Pro Trp Ser Ser Met Ile Pro His Glu His
 145 150 155 160
 Ser Phe Leu Pro Ser Ser Tyr Arg Gly Lys Asp Leu Gln Glu Asn Tyr
 165 170 175
 Cys Arg Asn Pro Arg Gly Glu Glu Gly Gly Pro Trp Cys Phe Thr Ser
 180 185 190
 Asn Pro Glu Val Arg Tyr Glu Val Cys Asp Ile Pro Gln Cys Ser Glu
 195 200 205
 Val Glu Cys Met Thr Cys Asn Gly Glu Ser Tyr Arg Gly Leu Met Asp
 210 215 220
 His Thr Glu Ser Gly Lys Ile Cys Gln Arg Trp Asp His Gln Thr Pro
 225 230 235 240
 His Arg His Lys Phe Leu Pro Glu Arg Tyr Pro Asp Lys Gly Phe Asp
 245 250 255
 Asp Asn Tyr Cys Arg Asn Pro Asp Gly Gln Pro Arg Pro Trp Cys Tyr
 260 265 270
 Thr Leu Asp Pro His Thr Arg Trp Glu Tyr Cys Ala Ile Lys Thr Cys
 275 280 285
 Ala Asp Asn Thr Met Asn Asp Thr Asp Val Pro Leu Glu Thr Thr Glu
 290 295 300
 Cys Ile Gln Gly Gln Gly Glu Gly Tyr Arg Gly Thr Val Asn Thr Ile
 Page 20

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

305      310      315      320
Trp Asn Gly Ile Pro Cys Gln Arg Trp Asp Ser Gln Tyr Pro His Glu
325      330      335
His Asp Met Thr Pro Glu Asn Phe Lys Cys Lys Asp Leu Arg Glu Asn
340      345      350
Tyr Cys Arg Asn Pro Asp Gly Ser Glu Ser Pro Trp Cys Phe Thr Thr
355      360      365
Asp Pro Asn Ile Arg Val Gly Tyr Cys Ser Gln Ile Pro Asn Cys Asp
370      375      380
Met Ser His Gly Gln Asp Cys Tyr Arg Gly Asn Gly Lys Asn Tyr Met
385      390      395      400
Gly Asn Leu Ser Gln Thr Arg Ser Gly Leu Thr Cys Ser Met Trp Asp
405      410      415
Lys Asn Met Glu Asp Leu His Arg His Ile Phe Trp Glu Pro Asp Ala
420      425      430
Ser Lys Leu Asn Glu Asn Tyr Cys Arg Asn Pro Asp Asp Asp Ala His
435      440      445
Gly Pro Trp Cys Tyr Thr Gly Asn Pro Leu Ile Pro Trp Asp Tyr Cys
450      455      460
Pro Ile Ser Arg Cys Glu Gly Asp Thr Thr Pro Thr Ile Val Asn Leu
465      470      475      480
Asp His Pro Val Ile Ser Cys Ala Lys Thr Lys Gln Leu Arg Val Val
485      490      495
Asn Gly Ile Pro Thr Arg Thr Asn Ile Gly Trp Met Val Ser Leu Arg
500      505      510
Tyr Arg Asn Lys His Ile Cys Gly Gly Ser Leu Ile Lys Glu Ser Trp
515      520      525
Val Leu Thr Ala Arg Gln Cys Phe Pro Ser Arg Asp Leu Lys Asp Tyr
530      535      540
Glu Ala Trp Leu Gly Ile His Asp Val His Gly Arg Gly Asp Glu Lys
545      550      555      560
Cys Lys Gln Val Leu Asn Val Ser Gln Leu Val Tyr Gly Pro Glu Gly
565      570      575
Ser Asp Leu Val Leu Met Lys Leu Ala Arg Pro Ala Val Leu Asp Asp
580      585      590
Phe Val Ser Thr Ile Asp Leu Pro Asn Tyr Gly Cys Thr Ile Pro Glu
595      600      605
Lys Thr Ser Cys Ser Val Tyr Gly Trp Gly Tyr Thr Gly Leu Ile Asn
610      615      620
Tyr Asp Gly Leu Leu Arg Val Ala His Leu Tyr Ile Met Gly Asn Glu
625      630      635      640
Lys Cys Ser Gln His His Arg Gly Lys Val Thr Leu Asn Glu Ser Glu
645      650      655
Ile Cys Ala Gly Ala Glu Lys Ile Gly Ser Gly Pro Cys Glu Gly Asp
660      665      670
Tyr Gly Gly Pro Leu Val Cys Glu Gln His Lys Met Arg Met Val Leu
675      680      685
Gly Val Ile Val Pro Gly Arg Gly Cys Ala Ile Pro Asn Arg Pro Gly
690      695      700
Ile Phe Val Arg Val Ala Tyr Tyr Ala Lys Trp Ile His Lys Ile Ile
705      710      715      720
Leu Thr Tyr Lys Val Pro Gln Ser
725

```

<210> 33
 <211> 232
 <212> PRT
 <213> Homo sapiens

<400> 33
 Met Asn Phe Leu Leu Ser Trp Val His Trp Ser Leu Ala Leu Leu Leu
 1 5 10 15
 Tyr Leu His His Ala Lys Trp Ser Gln Ala Ala Pro Met Ala Glu Gly
 20 25 30
 Gly Gly Gln Asn His His Glu Val Lys Phe Met Asp Val Tyr Gln
 35 40 45
 Arg Ser Tyr Cys His Pro Ile Glu Thr Leu Val Asp Ile Phe Gln Glu

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

50          55          60
Tyr Pro Asp Glu Ile Glu Tyr Ile Phe Lys Pro Ser Cys Val Pro Leu
65 Met Arg Cys Gly Gly Cys Cys Asn Asp Glu Gly Leu Glu Cys Val Pro
85 Thr Glu Glu Ser Asn Ile Thr Met Gln Ile Met Arg Ile Lys Pro His
100 Gln Gly Gln His Ile Gly Glu Met Ser Phe Leu Gln His Asn Lys Cys
115 Glu Cys Arg Pro Lys Lys Asp Arg Ala Arg Gln Glu Lys Lys Ser Val
130 Arg Gly Lys Gly Lys Gly Gln Lys Arg Lys Arg Lys Lys Ser Arg Tyr
145 Lys Ser Trp Ser Val Tyr Val Gly Ala Arg Cys Cys Leu Met Pro Trp
165 Ser Leu Pro Gly Pro His Pro Cys Gly Pro Cys Ser Glu Arg Arg Lys
180 His Leu Phe Val Gln Asp Pro Gln Thr Cys Lys Cys Ser Cys Lys Asn
195 Thr Asp Ser Arg Cys Lys Ala Arg Gln Leu Glu Leu Asn Glu Arg Thr
210 Cys Arg Cys Asp Lys Pro Arg Arg
225 230

```

```

<210> 34
<211> 240
<212> PRT
<213> Homo sapiens

```

```

<400> 34
Met Leu Ala Leu Leu Cys Ser Cys Leu Leu Leu Ala Ala Gly Ala Ser
1 5 10 15
Asp Ala Trp Thr Gly Glu Asp Ser Ala Glu Pro Asn Ser Asp Ser Ala
20 25 30
Glu Trp Ile Arg Asp Met Tyr Ala Lys Val Thr Glu Ile Trp Gln Glu
35 40 45
Val Met Gln Arg Arg Asp Asp Asp Gly Ala Leu His Ala Ala Cys Gln
50 55 60
Val Gln Pro Ser Ala Thr Leu Asp Ala Ala Gln Pro Arg Val Thr Gly
65 70 75 80
Val Val Leu Phe Arg Gln Leu Ala Pro Arg Ala Lys Leu Asp Ala Phe
85 90 95
Phe Ala Leu Glu Gly Phe Pro Thr Glu Pro Asn Ser Ser Ser Arg Ala
100 105 110
Ile His Val His Gln Phe Gly Asp Leu Ser Gln Gly Cys Glu Ser Thr
115 120 125
Gly Pro His Tyr Asn Pro Leu Ala Val Pro His Pro Gln His Pro Gly
130 135 140
Asp Phe Gly Asn Phe Ala Val Arg Asp Gly Ser Leu Trp Arg Tyr Arg
145 150 155 160
Ala Gly Leu Ala Ala Ser Leu Ala Gly Pro His Ser Ile Val Gly Arg
165 170 175
Ala Val Val Val His Ala Gly Glu Asp Leu Gly Arg Gly Gly Asn
180 185 190
Gln Ala Ser Val Glu Asn Gly Asn Ala Gly Arg Arg Leu Ala Cys Cys
195 200 205
Val Val Gly Val Cys Gly Pro Gly Leu Trp Glu Arg Gln Ala Arg Glu
210 215 220
His Ser Glu Arg Lys Lys Arg Arg Arg Glu Ser Glu Cys Lys Ala Ala
225 230 235 240

```

```

<210> 35
<211> 120
<212> PRT
<213> Homo sapiens

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

<400> 35

```

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

```

<210> 36

<211> 328

<212> PRT

<213> Homo sapiens

<400> 36

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

<210> 37
 <211> 92
 <212> PRT
 <213> Homo sapiens

<400> 37
 Met Ser Glu Leu Glu Lys Ala Met Val Ala Leu Ile Asp Val Phe His
 1 5 10 15
 Gln Tyr Ser Gly Arg Glu Gly Asp Lys His Lys Leu Lys Lys Ser Glu
 20 25 30
 Leu Lys Glu Leu Ile Asn Asn Glu Leu Ser His Phe Leu Glu Glu Ile
 35 40 45
 Lys Glu Gln Glu Val Val Asp Lys Val Met Glu Thr Leu Asp Asn Asp
 50 55 60
 Gly Asp Gly Glu Cys Asp Phe Gln Glu Phe Met Ala Phe Val Ala Met
 65 70 75 80
 Val Thr Thr Ala Cys His Glu Phe Phe Glu His Glu
 85 90

<210> 38
 <211> 174
 <212> PRT
 <213> Homo sapiens

<400> 38
 Met Ser Ser Gln Ile Arg Gln Asn Tyr Ser Thr Asp Val Glu Ala Ala
 1 5 10 15
 Val Asn Ser Leu Val Asn Leu Tyr Leu Gln Ala Ser Tyr Thr Tyr Leu
 20 25 30
 Ser Leu Gly Phe Tyr Phe Asp Arg Asp Asp Val Ala Leu Glu Gly Val
 35 40 45
 Ser His Phe Phe Arg Glu Leu Ala Glu Glu Lys Arg Glu Gly Tyr Glu
 50 55 60
 Arg Leu Leu Lys Met Gln Asn Gln Arg Gly Gly Arg Ala Leu Phe Gln
 65 70 75 80
 Asp Ile Lys Lys Pro Ala Glu Asp Glu Trp Gly Lys Thr Pro Asp Ala
 85 90 95
 Met Lys Ala Ala Met Ala Leu Glu Lys Lys Leu Asn Gln Ala Leu Leu
 100 105 110
 Asp Leu His Ala Leu Gly Ser Ala Arg Thr Asp Pro His Leu Cys Asp
 115 120 125
 Phe Leu Glu Thr His Phe Leu Asp Glu Glu Val Lys Leu Ile Lys Lys
 130 135 140
 Met Gly Asp His Leu Thr Asn Leu His Arg Leu Gly Gly Pro Glu Ala
 145 150 155 160
 Gly Leu Gly Glu Tyr Leu Phe Glu Arg Leu Thr Leu Lys His
 165 170

<210> 39
 <211> 217
 <212> PRT
 <213> Homo sapiens

<400> 39
 Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
 1 5 10 15
 Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
 20 25 30
 Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln
 35 40 45
 Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys
 50 55 60
 Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe
 65 70 75 80
 Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys
 Page 24

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

      85      90      95
Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Leu Ile Gln Ser Trp
100 105 110
Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val
115 120 125
Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu
130 135 140
Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg
145 150 155 160
Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser
165 170 175
His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe
180 185 190
Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys
195 200 205
Arg Ser Val Glu Gly Ser Cys Gly Phe
210 215

```

<210> 40

<211> 381

<212> PRT

<213> Homo sapiens

<400> 40

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

Val Val Asp 340 Gly Val Lys Leu Leu 345 Ile Glu Met Glu Gln 350 Arg Leu Glu
 355
 Gln Gly Gln Ala Ile Asp Asp 360 Leu Met Pro Ala Gln 365 Lys
 370 375 380

<210> 41
 <211> 540
 <212> PRT
 <213> Homo sapiens

<400> 41
 Met Leu Arg Leu Pro Thr Val Phe Arg Gln Met Arg Pro Val Ser Arg
 1 5 10 15
 Val Leu Ala Pro His Leu Thr Arg Ala Tyr Ala Lys Asp Val Lys Phe
 20 25 30
 Gly Ala Asp Ala Arg Ala Leu Met Leu Gln Gly Val Asp Leu Leu Ala
 35 40 45
 Asp Ala Val Ala Val Thr Met Gly Pro Lys Gly Arg Thr Val Ile Ile
 50 55 60
 Glu Gln Ser Trp Gly Ser Pro Lys Val Thr Lys Asp Gly Val Thr Val
 65 70 75 80
 Ala Lys Ser Ile Asp Leu Lys Asp Lys Tyr Lys Asn Ile Gly Ala Lys
 85 90 95
 Leu Val Gln Asp Val Ala Asn Asn Thr Asn Glu Glu Ala Gly Asp Gly
 100 105 110
 Thr Thr Thr Ala Thr Val Leu Ala Arg Ser Ile Ala Lys Glu Gly Phe
 115 120 125
 Glu Lys Ile Ser Lys Gly Ala Asn Pro Val Glu Ile Arg Arg Gly Val
 130 135 140
 Met Leu Ala Val Asp Ala Val Ile Ala Glu Leu Lys Lys Gln Ser Lys
 145 150 155 160
 Pro Val Thr Thr Pro Glu Glu Ile Ala Gln Val Ala Thr Ile Ser Ala
 165 170 175
 Asn Gly Asp Lys Glu Ile Gly Asn Ile Ile Ser Asp Ala Met Lys Lys
 180 185 190
 Val Gly Arg Lys Gly Val Ile Thr Val Lys Asp Gly Lys Thr Leu Asn
 195 200 205
 Asp Glu Leu Glu Ile Ile Glu Gly Met Lys Phe Asp Arg Gly Tyr Ile
 210 215 220
 Ser Pro Tyr Phe Ile Asn Thr Ser Lys Gly Gln Lys Cys Glu Phe Gln
 225 230 235 240
 Asp Ala Tyr Val Leu Leu Ser Glu Lys Lys Ile Ser Ser Ile Gln Ser
 245 250 255
 Ile Val Pro Ala Leu Glu Ile Ala Asn Ala His Arg Lys Pro Leu Val
 260 265 270
 Ile Ile Ala Glu Asp Val Asp Gly Glu Ala Leu Ser Thr Leu Val Leu
 275 280 285
 Asn Arg Leu Lys Val Gly Leu Gln Val Val Ala Val Lys Ala Pro Gly
 290 295 300
 Phe Gly Asp Asn Arg Lys Asn Gln Leu Lys Asp Met Ala Ile Ala Thr
 305 310 315 320
 Gly Gly Ala Val Phe Gly Glu Glu Gly Leu Thr Leu Asn Leu Glu Asp
 325 330 335
 Val Gln Pro His Asp Leu Gly Lys Val Gly Glu Val Ile Val Thr Lys
 340 345 350
 Asp Asp Ala Met Leu Leu Lys Gly Lys Gly Asp Lys Ala Gln Ile Glu
 355 360 365
 Lys Arg Ile Gln Glu Ile Ile Glu Gln Leu Asp Val Thr Thr Ser Glu
 370 375 380
 Tyr Glu Lys Glu Lys Leu Asn Glu Arg Leu Ala Lys Leu Ser Asp Gly
 385 390 395 400
 Val Ala Val Leu Lys Val Gly Gly Thr Ser Asp Val Glu Val Asn Glu
 405 410 415
 Lys Lys Asp Arg Val Thr Asp Ala Leu Asn Ala Thr Arg Ala Ala Val
 420 425 430
 Glu Glu Gly Ile Val Leu Gly Gly Gly Cys Ala Leu Leu Arg Cys Ile

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

      435      440      445
Pro Ala Leu Asp Ser Leu Thr Pro Ala Asn Glu Asp Gln Lys Ile Gly
  450      455      460
Ile Glu Ile Ile Lys Arg Thr Leu Lys Ile Pro Ala Met Thr Ile Ala
  465      470      475      480
Lys Asn Ala Gly Val Glu Gly Ser Leu Ile Val Glu Lys Ile Met Gln
      485      490      495
Ser Ser Ser Glu Val Gly Tyr Asp Ala Met Ala Gly Asp Phe Val Asn
      500      505      510
Met Val Glu Lys Gly Ile Ile Asp Pro Thr Lys Val Val Arg Thr Ala
      515      520      525
Leu Leu Asp Ala Ala Gly Val Ala Ser Leu Leu Thr
      530      535      540

```

<210> 42
 <211> 92
 <212> PRT
 <213> Homo sapiens

```

<400> 42
Met Lys Leu Cys Val Thr Val Leu Ser Leu Leu Met Leu Val Ala Ala
  1      5      10      15
Phe Cys Ser Pro Ala Leu Ser Ala Pro Met Gly Ser Asp Pro Pro Thr
      20      25      30
Ala Cys Cys Phe Ser Tyr Thr Ala Arg Lys Leu Pro Arg Asn Phe Val
      35      40      45
Val Asp Tyr Tyr Glu Thr Ser Ser Leu Cys Ser Gln Pro Ala Val Val
      50      55      60
Phe Gln Thr Lys Arg Ser Lys Gln Val Cys Ala Asp Pro Ser Glu Ser
  65      70      75      80
Trp Val Gln Glu Tyr Val Tyr Asp Leu Glu Leu Asn
      85      90

```

<210> 43
 <211> 207
 <212> PRT
 <213> Homo sapiens

```

<400> 43
Met Ala Pro Phe Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp
  1      5      10      15
Leu Ile Ala Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln
      20      25      30
Thr Ala Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly
      35      40      45
Thr Pro Glu Val Asn Gln Thr Leu Tyr Gln Arg Tyr Glu Ile Lys
      50      55      60
Met Thr Lys Met Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala Asp
  65      70      75      80
Ile Arg Phe Val Tyr Thr Pro Ala Met Glu Ser Val Cys Gly Tyr Phe
      85      90      95
His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala Gly Lys Leu
      100      105      110
Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe Val Ala Pro Trp
      115      120      125
Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe Thr Lys Thr Tyr Thr
      130      135      140
Val Gly Cys Glu Glu Cys Thr Val Phe Pro Cys Leu Ser Ile Pro Cys
  145      150      155      160
Lys Leu Gln Ser Gly Thr His Cys Leu Trp Thr Asp Gln Leu Leu Gln
      165      170      175
Gly Ser Glu Lys Gly Phe Gln Ser Arg His Leu Ala Cys Leu Pro Arg
      180      185      190
Glu Pro Gly Leu Cys Thr Trp Gln Ser Leu Arg Ser Gln Ile Ala
      195      200      205

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

<210> 44
 <211> 745
 <212> PRT
 <213> Homo sapiens

<400> 44

```

Met Gly Val Pro Phe Phe Ser Ser Leu Arg Cys Met Val Asp Leu Gly
 1      5      10      15
Pro Cys Trp Ala Gly Gly Leu Thr Ala Glu Met Lys Leu Leu Leu Ala
 20      25      30
Leu Ala Gly Leu Leu Ala Ile Leu Ala Thr Pro Gln Pro Ser Glu Gly
 35      40      45
Ala Ala Pro Ala Val Leu Gly Glu Val Asp Thr Ser Leu Val Leu Ser
 50      55      60
Ser Met Glu Glu Ala Lys Gln Leu Val Asp Lys Ala Tyr Lys Glu Arg
 65      70      75      80
Arg Glu Ser Ile Lys Gln Arg Leu Arg Ser Gly Ser Ala Ser Pro Met
 85      90      95
Glu Leu Leu Ser Tyr Phe Lys Gln Pro Val Ala Ala Thr Arg Thr Ala
 100      105      110
Val Arg Ala Ala Asp Tyr Leu His Val Ala Leu Asp Leu Leu Glu Arg
 115      120      125
Lys Leu Arg Ser Leu Trp Arg Arg Pro Phe Asn Val Thr Asp Val Leu
 130      135      140
Thr Pro Ala Gln Leu Asn Val Leu Ser Lys Ser Ser Gly Cys Ala Tyr
 145      150      155      160
Gln Asp Val Gly Val Thr Cys Pro Glu Gln Asp Lys Tyr Arg Thr Ile
 165      170      175
Thr Gly Met Cys Asn Asn Arg Arg Ser Pro Thr Leu Gly Ala Ser Asn
 180      185      190
Arg Ala Phe Val Arg Trp Leu Pro Ala Glu Tyr Glu Asp Gly Phe Ser
 195      200      205
Leu Pro Tyr Gly Trp Thr Pro Gly Val Lys Arg Asn Gly Phe Pro Val
 210      215      220
Ala Leu Ala Arg Ala Val Ser Asn Glu Ile Val Arg Phe Pro Thr Asp
 225      230      235      240
Gln Leu Thr Pro Asp Gln Glu Arg Ser Leu Met Phe Met Gln Trp Gly
 245      250      255
Gln Leu Leu Asp His Asp Leu Asp Phe Thr Pro Glu Pro Ala Ala Arg
 260      265      270
Ala Ser Phe Val Thr Gly Val Asn Cys Glu Thr Ser Cys Val Gln Gln
 275      280      285
Pro Pro Cys Phe Pro Leu Lys Ile Pro Pro Asn Asp Pro Arg Ile Lys
 290      295      300
Asn Gln Ala Asp Cys Ile Pro Phe Phe Arg Ser Cys Pro Ala Cys Pro
 305      310      315      320
Gly Ser Asn Ile Thr Ile Arg Asn Gln Ile Asn Ala Leu Thr Ser Phe
 325      330      335
Val Asp Ala Ser Met Val Tyr Gly Ser Glu Glu Pro Leu Ala Arg Asn
 340      345      350
Leu Arg Asn Met Ser Asn Gln Leu Gly Leu Leu Ala Val Asn Gln Arg
 355      360      365
Phe Gln Asp Asn Gly Arg Ala Leu Leu Pro Phe Asp Asn Leu His Asp
 370      375      380
Asp Pro Cys Leu Leu Thr Asn Arg Ser Ala Arg Ile Pro Cys Phe Leu
 385      390      395      400
Ala Gly Asp Thr Arg Ser Ser Glu Met Pro Glu Leu Thr Ser Met His
 405      410      415
Thr Leu Leu Leu Arg Glu His Asn Arg Leu Ala Thr Glu Leu Lys Ser
 420      425      430
Leu Asn Pro Arg Trp Asp Gly Glu Arg Leu Tyr Gln Glu Ala Arg Lys
 435      440      445
Ile Val Gly Ala Met Val Gln Ile Ile Thr Tyr Arg Asp Tyr Leu Pro
 450      455      460
Leu Val Leu Gly Pro Thr Ala Met Arg Lys Tyr Leu Pro Thr Tyr Arg
 465      470      475      480
Ser Tyr Asn Asp Ser Val Asp Pro Arg Ile Ala Asn Val Phe Thr Asn

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

      485      490      495
Ala Phe Arg Tyr Gly His Thr Leu Ile Gln Pro Phe Met Phe Arg Leu
      500      505      510
Asp Asn Arg Tyr Gln Pro Met Glu Pro Asn Pro Arg Val Pro Leu Ser
      515      520      525
Arg Val Phe Phe Ala Ser Trp Arg Val Val Leu Glu Gly Gly Ile Asp
      530      535      540
Pro Ile Leu Arg Gly Leu Met Ala Thr Pro Ala Lys Leu Asn Arg Gln
      545      550      555      560
Asn Gln Ile Ala Val Asp Glu Ile Arg Glu Arg Leu Phe Glu Gln Val
      565      570      575
Met Arg Ile Gly Leu Asp Leu Pro Ala Leu Asn Met Gln Arg Ser Arg
      580      585      590
Asp His Gly Leu Pro Gly Tyr Asn Ala Trp Arg Arg Phe Cys Gly Leu
      595      600      605
Pro Gln Pro Glu Thr Val Gly Gln Leu Gly Thr Val Leu Arg Asn Leu
      610      615      620
Lys Leu Ala Arg Lys Leu Met Glu Gln Tyr Gly Thr Pro Asn Asn Ile
      625      630      635      640
Asp Ile Trp Met Gly Gly Val Ser Glu Pro Leu Lys Arg Lys Gly Arg
      645      650      655
Val Gly Pro Leu Leu Ala Cys Ile Ile Gly Thr Gln Phe Arg Lys Leu
      660      665      670
Arg Asp Gly Asp Arg Phe Trp Trp Glu Asn Glu Gly Val Phe Ser Met
      675      680      685
Gln Gln Arg Gln Ala Leu Ala Gln Ile Ser Leu Pro Arg Ile Ile Cys
      690      695      700
Asp Asn Thr Gly Ile Thr Thr Val Ser Lys Asn Asn Ile Phe Met Ser
      705      710      715      720
Asn Ser Tyr Pro Arg Asp Phe Val Asn Cys Ser Thr Leu Pro Ala Leu
      725      730      735
Asn Leu Ala Ser Trp Arg Glu Ala Ser
      740      745

```

<210> 45
 <211> 110
 <212> PRT
 <213> Homo sapiens

```

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

```

<210> 46
 <211> 233
 <212> PRT
 <213> Homo sapiens

```

<400> 46
Met Ser Thr Glu Ser Met Ile Arg Asp Val Glu Leu Ala Glu Glu Ala
  1      5      10      15
Leu Pro Lys Lys Thr Gly Gly Pro Gln Gly Ser Arg Arg Cys Leu Phe
      20      25      30
Leu Ser Leu Phe Ser Phe Leu Ile Val Ala Gly Ala Thr Thr Leu Phe

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 47
 <211> 125
 <212> PRT
 <213> Homo sapiens

```

<400> 47
Met Lys Lys Ser Gly Val Leu Phe Leu Leu Gly Ile Ile Leu Leu Val
 1      5      10      15
Leu Ile Gly Val Gln Gly Thr Pro Val Val Arg Lys Gly Arg Cys Ser
      20      25      30
Cys Ile Ser Thr Asn Gln Gly Thr Ile His Leu Gln Ser Leu Lys Asp
      35      40      45
Leu Lys Gln Phe Ala Pro Ser Pro Ser Cys Glu Lys Ile Glu Ile Ile
      50      55      60
Ala Thr Leu Lys Asn Gly Val Gln Thr Cys Leu Asn Pro Asp Ser Ala
 65      70      75      80
Asp Val Lys Glu Leu Ile Lys Lys Trp Glu Lys Gln Val Ser Gln Lys
      85      90      95
Lys Lys Gln Lys Asn Gly Lys Lys His Gln Lys Lys Lys Val Leu Lys
      100      105      110
Val Arg Lys Ser Gln Arg Ser Arg Gln Lys Lys Thr Thr
      115      120      125

```

<210> 48
 <211> 335
 <212> PRT
 <213> Homo sapiens

```

<400> 48
Met Leu Gly Ile Trp Thr Leu Leu Pro Leu Val Leu Thr Ser Val Ala
 1      5      10      15
Arg Leu Ser Ser Lys Ser Val Asn Ala Gln Val Thr Asp Ile Asn Ser
      20      25      30
Lys Gly Leu Glu Leu Arg Lys Thr Val Thr Thr Val Glu Thr Gln Asn
      35      40      45
Leu Glu Gly Leu His His Asp Gly Gln Phe Cys His Lys Pro Cys Pro
      50      55      60
Pro Gly Glu Arg Lys Ala Arg Asp Cys Thr Val Asn Gly Asp Glu Pro
 65      70      75      80
Asp Cys Val Pro Cys Gln Glu Gly Lys Glu Tyr Thr Asp Lys Ala His

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

				85				90					95				
Phe	Ser	Ser	Lys	Cys	Arg	Arg	Cys	Arg	Leu	Cys	Asp	Glu	Gly	His	Gly		
			100					105					110				
Leu	Glu	Val	Glu	Ile	Asn	Cys	Thr	Arg	Thr	Gln	Asn	Thr	Lys	Cys	Arg		
			115					120					125				
Cys	Lys	Pro	Asn	Phe	Phe	Cys	Asn	Ser	Thr	Val	Cys	Glu	His	Cys	Asp		
			130				135				140						
Pro	Cys	Thr	Lys	Cys	Glu	His	Gly	Ile	Ile	Lys	Glu	Cys	Thr	Leu	Thr		
145					150					155					160		
Ser	Asn	Thr	Lys	Cys	Lys	Glu	Glu	Gly	Ser	Arg	Ser	Asn	Leu	Gly	Trp		
					165				170					175			
Leu	Cys	Leu	Leu	Leu	Leu	Pro	Ile	Pro	Leu	Ile	Val	Trp	Val	Lys	Arg		
					180			185					190				
Lys	Glu	Val	Gln	Lys	Thr	Cys	Arg	Lys	His	Arg	Lys	Glu	Asn	Gln	Gly		
			195				200					205					
Ser	His	Glu	Ser	Pro	Thr	Leu	Asn	Pro	Glu	Thr	Val	Ala	Ile	Asn	Leu		
						215					220						
Ser	Asp	Val	Asp	Leu	Ser	Lys	Tyr	Ile	Thr	Thr	Ile	Ala	Gly	Val	Met		
225					230				235					240			
Thr	Leu	Ser	Gln	Val	Lys	Gly	Phe	Val	Arg	Lys	Asn	Gly	Val	Asn	Glu		
				245				250					255				
Ala	Lys	Ile	Asp	Glu	Ile	Lys	Asn	Asp	Asn	Val	Gln	Asp	Thr	Ala	Glu		
			260				265					270					
Gln	Lys	Val	Gln	Leu	Leu	Arg	Asn	Trp	His	Gln	Leu	His	Gly	Lys	Lys		
			275				280					285					
Glu	Ala	Tyr	Asp	Thr	Leu	Ile	Lys	Asp	Leu	Lys	Lys	Ala	Asn	Leu	Cys		
			290			295					300						
Thr	Leu	Ala	Glu	Lys	Ile	Gln	Thr	Ile	Ile	Leu	Lys	Asp	Ile	Thr	Ser		
305					310					315					320		
Asp	Ser	Glu	Asn	Ser	Asn	Phe	Arg	Asn	Glu	Ile	Gln	Ser	Leu	Val			
				325					330					335			

<210> 49

<211> 406

<212> PRT

<213> Homo sapiens

<400> 49

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

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

225      230      235      240
Asn Tyr Ser Gln Val Asp Ile Gly Leu Ile Lys Leu Lys Gln Lys Val
      245      250      255
Ser Val Asn Glu Arg Val Met Pro Ile Cys Leu Pro Ser Lys Asp Tyr
      260      265      270
Ala Glu Val Gly Arg Val Gly Tyr Val Ser Gly Trp Gly Arg Asn Ala
      275      280      285
Asn Phe Lys Phe Thr Asp His Leu Lys Tyr Val Met Leu Pro Val Ala
      290      295      300
Asp Gln Asp Gln Cys Ile Arg His Tyr Glu Gly Ser Thr Val Pro Glu
305      310      315      320
Lys Lys Thr Pro Lys Ser Pro Val Gly Val Gln Pro Ile Leu Asn Glu
      325      330      335
His Thr Phe Cys Ala Gly Met Ser Lys Tyr Gln Glu Asp Thr Cys Tyr
      340      345      350
Gly Asp Ala Gly Ser Ala Phe Ala Val His Asp Leu Glu Glu Asp Thr
      355      360      365
Trp Tyr Ala Thr Gly Ile Leu Ser Phe Asp Lys Ser Cys Ala Val Ala
370      375      380
Glu Tyr Gly Val Tyr Val Lys Val Thr Ser Ile Gln Asp Trp Val Gln
385      390      395      400
Lys Thr Ile Ala Glu Asn
      405

```

<210> 50
 <211> 440
 <212> PRT
 <213> Homo sapiens

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

290          295          300
Pro Gly Glu Ser Glu His Leu Leu Glu Pro Ala Glu Ala Glu Arg Ser
305 Gln Arg Arg Arg Leu Leu Val Pro Ala Asn Glu Gly Asp Pro Thr Glu
325 Phe Asp Asp Phe Ala Asp Leu Val Pro Phe Asp
340 Leu Met Arg Lys Leu Gly Leu Met Asp Asn Glu Ile
355 Val Ala Lys Ala Glu Ala Ala Gly His Arg Asp Thr Leu Tyr Thr
370 Met Leu Ile Lys Trp Val Asn Lys Thr Gly Arg Asp Ala Ser Val His
385 Thr Leu Leu Asp Ala Leu Glu Thr Leu Gly Glu Arg Leu Ala Lys Gln
405 Lys Ile Glu Asp His Leu Leu Ser Ser Gly Lys Phe Met Tyr Leu Glu
420 Gly Asn Ala Asp Ser Ala Met Ser
435          440

```

<210> 51
 <211> 177
 <212> PRT
 <213> Homo sapiens

```

<400> 51
Met Arg Glu Arg Pro Arg Leu Gly Glu Asp Ser Ser Leu Ile Ser Leu
1 5 10 15
Phe Leu Gln Val Ala Phe Leu Ala Met Val Met Gly Thr His Thr
20 25 30
Tyr Ser His Trp Pro Ser Cys Cys Pro Ser Lys Gly Gln Asp Thr Ser
35 40 45
Glu Glu Leu Leu Arg Trp Ser Thr Val Pro Val Pro Pro Leu Glu Pro
50 55 60
Ala Arg Pro Asn Arg His Pro Glu Ser Cys Arg Ala Ser Glu Asp Gly
65 70 75 80
Pro Leu Asn Ser Arg Ala Ile Ser Pro Trp Arg Tyr Glu Leu Asp Arg
85 90 95
Asp Leu Asn Arg Leu Pro Gln Asp Leu Tyr His Ala Arg Cys Leu Cys
100 105 110
Pro His Cys Val Ser Leu Gln Thr Gly Ser His Met Asp Pro Arg Gly
115 120 125
Asn Ser Glu Leu Leu Tyr His Asn Gln Thr Val Phe Tyr Arg Arg Pro
130 135 140
Cys His Gly Glu Lys Gly Thr His Lys Gly Tyr Cys Leu Glu Arg Arg
145 150 155 160
Leu Tyr Arg Val Ser Leu Ala Cys Val Cys Val Arg Pro Arg Val Met
165 170 175
Gly

```

<210> 52
 <211> 267
 <212> PRT
 <213> Homo sapiens

```

<400> 52
Met Arg Leu Thr Val Leu Cys Ala Val Cys Leu Leu Pro Gly Ser Leu
1 5 10 15
Ala Leu Pro Leu Pro Gln Glu Ala Gly Gly Met Ser Glu Leu Gln Trp
20 25 30
Glu Gln Ala Gln Asp Tyr Leu Lys Arg Phe Tyr Leu Tyr Asp Ser Glu
35 40 45
Thr Lys Asn Ala Asn Ser Leu Glu Ala Lys Leu Lys Glu Met Gln Lys
50 55 60
Phe Phe Gly Leu Pro Ile Thr Gly Met Leu Asn Ser Arg Val Ile Glu
Page 33

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 53
 <211> 178
 <212> PRT
 <213> Homo sapiens

```

<400> 53
Met His Ser Ser Ala Leu Leu Cys Cys Leu Val Leu Leu Thr Gly Val
1      5      10      15
Arg Ala Ser Pro Gly Gln Gly Thr Gln Ser Glu Asn Ser Cys Thr His
      20      25      30
Phe Pro Gly Asn Leu Pro Asn Met Leu Arg Asp Leu Arg Asp Ala Phe
      35      40      45
Ser Arg Val Lys Thr Phe Phe Gln Met Lys Asp Gln Leu Asp Asn Leu
      50      55      60
Leu Leu Lys Glu Ser Leu Leu Glu Asp Phe Lys Gly Tyr Leu Gly Cys
      65      70      75
Gln Ala Leu Ser Glu Met Ile Gln Phe Tyr Leu Glu Glu Val Met Pro
      80      85      90
Gln Ala Glu Asn Gln Asp Pro Asp Ile Lys Ala His Val Asn Ser Leu
      95      100      105
Gly Glu Asn Leu Lys Thr Leu Arg Leu Arg Arg Cys His Arg
      110      115      120
Phe Leu Pro Cys Glu Asn Lys Ser Lys Ala Val Glu Gln Val Lys Asn
      125      130      135
Ala Phe Asn Lys Leu Gln Glu Lys Gly Ile Tyr Lys Ala Met Ser Glu
      140      145      150
Phe Asp Ile Phe Ile Asn Tyr Ile Glu Ala Tyr Met Thr Met Lys Ile
      155      160      165
      170      175
Arg Asn

```

<210> 54
 <211> 1332
 <212> PRT
 <213> Homo sapiens

```

<400> 54
Met Glu Ser His Ser Arg Ala Gly Lys Ser Arg Lys Ser Ala Lys Phe
1      5      10      15
Arg Ser Ile Ser Arg Ser Leu Met Leu Cys Asn Ala Lys Thr Ser Asp
      20      25

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

			20				25				30			
Asp	Gly	Ser	Pro	Asp	Glu	Lys	Tyr	Pro	Asp	Pro	Phe	Glu	Ile	Ser
		35				40					45			
Leu	Ala	Gln	Gly	Lys	Glu	Gly	Ile	Phe	His	Ser	Ser	Val	Gln	Leu
	50				55					60				Ala
Asp	Thr	Ser	Glu	Ala	Gly	Pro	Ser	Ser	Val	Pro	Asp	Leu	Ala	Leu
65					70					75				Ala
Ser	Glu	Ala	Ala	Gln	Leu	Gln	Ala	Ala	Gly	Asn	Asp	Arg	Gly	Lys
			85						90					Thr
Cys	Arg	Arg	Ile	Phe	Phe	Met	Lys	Glu	Ser	Ser	Thr	Ala	Ser	Ser
			100					105						Arg
Glu	Lys	Pro	Gly	Lys	Leu	Glu	Ala	Gln	Ser	Ser	Asn	Phe	Leu	Phe
		115					120					125		Pro
Lys	Ala	Cys	His	Gln	Arg	Ala	Arg	Ser	Asn	Ser	Thr	Ser	Val	Asn
	130					135					140			Pro
Tyr	Cys	Thr	Arg	Glu	Ile	Asp	Phe	Pro	Met	Thr	Lys	Lys	Ser	Ala
145					150					155				Ala
Pro	Thr	Asp	Arg	Gln	Pro	Tyr	Ser	Leu	Cys	Ser	Asn	Arg	Lys	Ser
				165					170					Leu
Ser	Gln	Gln	Leu	Asp	Cys	Pro	Ala	Gly	Lys	Ala	Ala	Gly	Thr	Ser
			180					185					190	Arg
Pro	Thr	Arg	Ser	Leu	Ser	Thr	Ala	Gln	Leu	Val	Gln	Pro	Ser	Gly
		195					200					205		Gly
Leu	Gln	Ala	Ser	Val	Ile	Ser	Asn	Ile	Val	Leu	Met	Lys	Gly	Gln
	210					215					220			Ala
Lys	Gly	Leu	Gly	Phe	Ser	Ile	Val	Gly	Gly	Lys	Asp	Ser	Ile	Tyr
225					230					235				Gly
Pro	Ile	Gly	Ile	Tyr	Val	Lys	Thr	Ile	Phe	Ala	Gly	Gly	Ala	Ala
				245					250					Ala
Ala	Asp	Gly	Arg	Leu	Gln	Glu	Gly	Asp	Glu	Ile	Leu	Glu	Leu	Asn
			260					265					270	Gly
Glu	Ser	Met	Ala	Gly	Leu	Thr	His	Gln	Asp	Ala	Leu	Gln	Lys	Phe
		275					280					285		Lys
Gln	Ala	Lys	Lys	Gly	Leu	Leu	Thr	Leu	Thr	Val	Arg	Thr	Arg	Leu
	290					295					300			Thr
Ala	Pro	Pro	Ser	Leu	Cys	Ser	His	Leu	Ser	Pro	Pro	Leu	Cys	Arg
305					310					315				Ser
Leu	Ser	Ser	Ser	Thr	Cys	Ile	Thr	Lys	Asp	Ser	Ser	Ser	Phe	Ala
				325					330				335	Leu
Glu	Ser	Pro	Ser	Ala	Pro	Ile	Ser	Thr	Ala	Lys	Pro	Asn	Tyr	Arg
				340					345				350	Ile
Met	Val	Glu	Val	Ser	Leu	Gln	Lys	Glu	Ala	Gly	Val	Gly	Leu	Gly
		355					360					365		Ile
Gly	Leu	Cys	Ser	Val	Pro	Tyr	Phe	Gln	Cys	Ile	Ser	Gly	Ile	Phe
	370					375					380			Val
His	Thr	Leu	Ser	Pro	Gly	Ser	Val	Ala	His	Leu	Asp	Gly	Arg	Leu
385					390					395				Arg
Cys	Gly	Asp	Glu	Ile	Val	Glu	Ile	Ser	Asp	Ser	Pro	Val	His	Cys
				405					410					Leu
Thr	Leu	Asn	Glu	Val	Tyr	Thr	Ile	Leu	Ser	His	Cys	Asp	Pro	Gly
			420					425					430	Pro
Val	Pro	Ile	Ile	Val	Ser	Arg	His	Pro	Asp	Pro	Gln	Val	Ser	Glu
		435					440					445		Gln
Gln	Leu	Lys	Glu	Ala	Val	Ala	Gln	Ala	Val	Glu	Asn	Thr	Lys	Phe
	450					455					460			Gly
Lys	Glu	Arg	His	Gln	Trp	Ser	Leu	Glu	Gly	Val	Lys	Arg	Leu	Glu
465					470					475				Ser
Ser	Trp	His	Gly	Arg	Pro	Thr	Leu	Glu	Lys	Glu	Arg	Glu	Lys	Asn
				485					490					Ser
Ala	Pro	Pro	His	Arg	Arg	Ala	Gln	Lys	Val	Met	Ile	Arg	Ser	Ser
			500					505					510	Ser
Asp	Ser	Ser	Tyr	Met	Ser	Gly	Ser	Pro	Gly	Gly	Ser	Pro	Gly	Ser
		515					520					525		Gly
Ser	Ala	Glu	Lys	Pro	Ser	Ser	Asp	Val	Asp	Ile	Ser	Thr	His	Ser
	530					535					540			Pro
Ser	Leu	Pro	Leu	Ala	Arg	Glu	Pro	Val	Val	Leu	Ser	Ile	Ala	Ser
545					550					555				Ser
Arg	Leu	Pro	Gln	Glu	Ser	Pro	Pro	Leu	Pro	Glu	Ser	Arg	Asp	Ser
														His

CEN5301WOPCTCORRECTEDSEQLIST.txt

Pro	Pro	Leu	Arg	565	Lys	Lys	Ser	Phe	570	Ile	Leu	Val	Arg	575	Pro
Met	Ser	Ser	Lys	580	Pro	Lys	Pro	Pro	585	Arg	Lys	Tyr	Phe	590	Asp
Ser	Asp	Pro	Gln	595	Lys	Ser	Leu	Glu	600	Arg	Glu	Asn	Ser	Ser	Cys
Ser	Gly	His	Thr	610	Pro	Pro	Thr	Cys	615	Gly	Gln	Glu	Ala	Arg	Leu
Pro	Leu	Leu	Leu	625	Pro	Gln	Glu	Asp	630	Thr	Ala	Gly	Arg	Ser	Pro
Ser	Ala	Gly	Cys	645	Pro	Gly	Pro	Gly	650	Ile	Gly	Pro	Gln	Thr	Lys
Thr	Glu	Gly	Glu	660	Pro	Gly	Trp	Arg	665	Arg	Ala	Ser	Pro	Val	Thr
Ser	Pro	Ile	Lys	675	His	Pro	Leu	Leu	680	Lys	Arg	Gln	Ala	Arg	Met
Ser	Phe	Asp	Thr	690	Thr	Ala	Glu	Asp	695	Pro	Trp	Val	Arg	Ile	Ser
Ile	Lys	Asn	Leu	705	Phe	Ser	Pro	Ile	710	Met	Ser	Glu	Asn	His	Gly
Pro	Leu	Gln	Pro	725	Asn	Ala	Ser	Leu	730	Asn	Glu	Glu	Glu	Gly	Thr
His	Pro	Asp	Gly	740	Thr	Pro	Pro	Lys	745	Leu	Asp	Thr	Ala	Asn	Gly
Lys	Val	Tyr	Lys	755	Ser	Ala	Asp	Ser	760	Ser	Thr	Val	Lys	Lys	Gly
Val	Ala	Pro	Lys	770	Pro	Ala	Trp	Phe	775	Arg	Gln	Ser	Leu	Lys	Gly
Asn	Arg	Ala	Ser	785	Asp	Pro	Arg	Gly	790	Leu	Pro	Asp	Pro	Ala	Leu
Gln	Pro	Ala	Pro	805	Ala	Ser	Arg	Glu	810	His	Leu	Gly	Ser	His	Ile
Ser	Ser	Ser	Ser	820	Ser	Ser	Ile	Arg	825	Gln	Arg	Ile	Ser	Ser	Phe
Phe	Gly	Ser	Ser	835	Gln	Leu	Pro	Asp	840	Lys	Gly	Ala	Gln	Arg	Leu
Gln	Pro	Ser	Ser	850	Gly	Glu	Ala	Ala	855	Lys	Pro	Leu	Gly	Lys	His
Gly	Arg	Phe	Ser	865	Gly	Leu	Leu	Gly	870	Arg	Gly	Ala	Ala	Pro	Thr
Pro	Gln	Gln	Pro	885	Gln	Val	Leu	Ser	890	Ser	Gly	Ser	Pro	Ala	Ala
Glu	Ala	Arg	Asp	900	Pro	Gly	Val	Ser	905	Glu	Ser	Pro	Pro	Pro	Gly
Pro	Asn	Gln	Lys	915	Thr	Leu	Pro	Pro	920	Gly	Pro	Asp	Pro	Leu	Leu
Leu	Ser	Thr	Gln	930	Ala	Glu	Pro	Ser	935	Gln	Gly	Pro	Val	Leu	Lys
Ser	Gln	Arg	Ala	945	Arg	Ser	Phe	Pro	950	Leu	Thr	Arg	Ser	Gln	Ser
Thr	Lys	Leu	Leu	965	Asp	Glu	Lys	Thr	970	Ser	Lys	Leu	Tyr	Ser	Ile
Gln	Val	Ser	Ser	980	Ala	Val	Met	Lys	985	Ser	Leu	Leu	Cys	Leu	Pro
Ile	Ser	Cys	Ala	995	Gln	Thr	Pro	Cys	1000	Ile	Pro	Lys	Glu	Gly	Ala
Thr	Ser	Ser	Ser	1010	Asn	Glu	Asp	Ser	1015	Ala	Ala	Asn	Gly	Ser	Ala
Ser	Ala	Leu	Asp	1025	Thr	Gly	Phe	Ser	1030	Leu	Asn	Leu	Ser	Glu	Leu
Tyr	Thr	Glu	Gly	1045	Leu	Thr	Glu	Ala	1050	Lys	Glu	Asp	Asp	Asp	Gly
Ser	Ser	Leu	Gln	1060	Ser	Gly	Gln	Ser	1065	Val	Ile	Ser	Leu	Leu	Ser
Glu	Leu	Lys	Lys	1075	Leu	Ile	Glu	Glu	1080	Val	Lys	Val	Leu	Asp	Glu
Leu	Lys	Gln	Leu	1090	Asp	Gly	Ile	His	1095	Val	Thr	Ile	Leu	His	Lys

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

1105      1110      1115      1120
Gly Ala Gly Leu Gly Phe Ser Leu Ala Gly Gly Ala Asp Leu Glu Asn
      1125      1130      1135
Lys Val Ile Thr Val His Arg Val Phe Pro Asn Gly Leu Ala Ser Gln
      1140      1145      1150
Glu Gly Thr Ile Gln Lys Gly Asn Glu Val Leu Ser Ile Asn Gly Lys
      1155      1160      1165
Ser Leu Lys Gly Thr Thr His His Asp Ala Leu Ala Ile Leu Arg Gln
      1170      1175      1180
Ala Arg Glu Pro Arg Gln Ala Val Ile Val Thr Arg Lys Leu Thr Pro
1185      1190      1195      1200
Glu Ala Met Pro Asp Leu Asn Ser Ser Thr Asp Ser Ala Ala Ser Ala
      1205      1210      1215
Ser Ala Ala Ser Asp Val Ser Val Glu Ser Thr Ala Glu Ala Thr Val
      1220      1225      1230
Cys Thr Val Thr Leu Glu Lys Met Ser Ala Gly Leu Gly Phe Ser Leu
      1235      1240      1245
Glu Gly Gly Lys Gly Ser Leu His Gly Asp Lys Pro Leu Thr Ile Asn
      1250      1255      1260
Arg Ile Phe Lys Gly Ala Ala Ser Glu Gln Ser Glu Thr Val Gln Pro
1265      1270      1275      1280
Gly Asp Glu Ile Leu Gln Leu Gly Gly Thr Ala Met Gln Gly Leu Thr
      1285      1290      1295
Arg Phe Glu Ala Trp Asn Ile Ile Lys Ala Leu Pro Asp Gly Pro Val
      1300      1305      1310
Thr Ile Val Ile Arg Arg Lys Ser Leu Gln Ser Lys Glu Thr Thr Ala
      1315      1320      1325
Ala Gly Asp Ser
1330

```

<210> 55
 <211> 109
 <212> PRT
 <213> Homo sapiens

```

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

```

<210> 56
 <211> 110
 <212> PRT
 <213> Homo sapiens

```

<400> 56
Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
 1      5      10      15
Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly
      20      25      30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
      35      40      45
Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
      50      55      60
Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

65					70					75					80
Ala	Leu	Glu	Gly	Ser	Leu	Gln	Lys	Arg	Gly	Ile	Val	Glu	Gln	Cys	Cys
				85					90					95	
Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu	Glu	Asn	Tyr	Cys	Asn		
			100					105					110		

<210> 57
 <211> 2813
 <212> PRT
 <213> Homo sapiens

<400> 57
 Met Ile Pro Ala Arg Phe Ala Gly Val Leu Leu Ala Leu Ala Leu Ile
 1 5 10 15
 Leu Pro Gly Thr Leu Cys Ala Glu Gly Thr Arg Gly Arg Ser Ser Thr
 20 25 30
 Ala Arg Cys Ser Leu Phe Gly Ser Asp Phe Val Asn Thr Phe Asp Gly
 35 40 45
 Ser Met Tyr Ser Phe Ala Gly Tyr Cys Ser Tyr Leu Leu Ala Gly Gly
 50 55 60
 Cys Gln Lys Arg Ser Phe Ser Ile Ile Gly Asp Phe Gln Asn Gly Lys
 65 70 75 80
 Arg Val Ser Leu Ser Val Tyr Leu Gly Glu Phe Phe Asp Ile His Leu
 85 90 95
 Phe Val Asn Gly Thr Val Thr Gln Gly Asp Gln Arg Val Ser Met Pro
 100 105 110
 Tyr Ala Ser Lys Gly Leu Tyr Leu Glu Thr Glu Ala Gly Tyr Tyr Lys
 115 120 125
 Leu Ser Gly Glu Ala Tyr Gly Phe Val Ala Arg Ile Asp Gly Ser Gly
 130 135 140
 Asn Phe Gln Val Leu Leu Ser Asp Arg Tyr Phe Asn Lys Thr Cys Gly
 145 150 155 160
 Leu Cys Gly Asn Phe Asn Ile Phe Ala Glu Asp Asp Phe Met Thr Gln
 165 170 175
 Glu Gly Thr Leu Thr Ser Asp Pro Tyr Asp Phe Ala Asn Ser Trp Ala
 180 185 190
 Leu Ser Ser Gly Glu Gln Trp Cys Glu Arg Ala Ser Pro Pro Ser Ser
 195 200 205
 Ser Cys Asn Ile Ser Ser Gly Glu Met Gln Lys Gly Leu Trp Glu Gln
 210 215 220
 Cys Gln Leu Leu Lys Ser Thr Ser Val Phe Ala Arg Cys His Pro Leu
 225 230 235 240
 Val Asp Pro Glu Pro Phe Val Ala Leu Cys Glu Lys Thr Leu Cys Glu
 245 250 255
 Cys Ala Gly Gly Leu Glu Cys Ala Cys Pro Ala Leu Leu Glu Tyr Ala
 260 265 270
 Arg Thr Cys Ala Gln Glu Gly Met Val Leu Tyr Gly Trp Thr Asp His
 275 280 285
 Ser Ala Cys Ser Pro Val Cys Pro Ala Gly Met Glu Tyr Arg Gln Cys
 290 295 300
 Val Ser Pro Cys Ala Arg Thr Cys Gln Ser Leu His Ile Asn Glu Met
 305 310 315 320
 Cys Gln Glu Arg Cys Val Asp Gly Cys Ser Cys Pro Glu Gly Gln Leu
 325 330 335
 Leu Asp Glu Gly Leu Cys Val Glu Ser Thr Glu Cys Pro Cys Val His
 340 345 350
 Ser Gly Lys Arg Tyr Pro Pro Gly Thr Ser Leu Ser Arg Asp Cys Asn
 355 360 365
 Thr Cys Ile Cys Arg Asn Ser Gln Trp Ile Cys Ser Asn Glu Glu Cys
 370 375 380
 Pro Gly Glu Cys Leu Val Thr Gly Gln Ser His Phe Lys Ser Phe Asp
 385 390 395 400
 Asn Arg Tyr Phe Thr Phe Ser Gly Ile Cys Gln Tyr Leu Leu Ala Arg
 405 410 415
 Asp Cys Gln Asp His Ser Phe Ser Ile Val Ile Glu Thr Val Gln Cys
 420 425 430
 Ala Asp Asp Arg Asp Ala Val Cys Thr Arg Ser Val Thr Val Arg Leu

CEN5301WOPCTCORRECTEDSEQLIST.txt

Pro	Gly	435	Leu	His	Asn	Ser	Leu	440	Val	Lys	Leu	Lys	His	445	Gly	Ala	Gly	Val
Ala	Met	450	Asp	Gly	Gln	Asp	Val	455	Gln	Leu	Pro	Leu	Leu	460	Lys	Gly	Asp	Leu
Arg	Ile	465	Gln	His	Thr	Val	Thr	470	Ala	Ser	Val	Arg	Leu	475	Ser	Tyr	Gly	Glu
Asp	Leu	485	Gln	Met	Asp	Trp	Asp	490	Gly	Arg	Gly	Arg	Leu	495	Leu	Val	Lys	Leu
Ser	Pro	500	Val	Tyr	Ala	Gly	Lys	505	Thr	Cys	Gly	Leu	Cys	510	Gly	Asn	Tyr	Asn
Gly	Asn	515	Gln	Gly	Asp	Asp	Phe	520	Leu	Thr	Pro	Ser	Gly	525	Leu	Ala	Glu	Pro
Arg	Val	530	Glu	Asp	Phe	Gly	Asn	535	Ala	Trp	Lys	Leu	His	540	Gly	Asp	Cys	Gln
Asp	Leu	545	Gln	Lys	Gln	His	Ser	550	Asp	Pro	Cys	Ala	Leu	555	Asn	Pro	Arg	Met
Thr	Arg	565	Phe	Ser	Glu	Glu	Ala	570	Cys	Ala	Val	Leu	Thr	575	Ser	Pro	Thr	Phe
Glu	Ala	580	Cys	His	Arg	Ala	Val	585	Ser	Pro	Leu	Pro	Tyr	590	Leu	Arg	Asn	Cys
Arg	Tyr	595	Asp	Val	Cys	Ser	Cys	600	Ser	Asp	Gly	Arg	Glu	605	Cys	Leu	Cys	Gly
Ala	Leu	610	Ala	Ser	Tyr	Ala	Ala	615	Ala	Cys	Ala	Gly	Arg	620	Gly	Val	Arg	Val
Ala	Trp	625	Arg	Glu	Pro	Gly	Arg	630	Cys	Glu	Leu	Asn	Cys	635	Pro	Lys	Gly	Gln
Val	Tyr	645	Leu	Gln	Cys	Gly	Thr	650	Pro	Cys	Asn	Leu	Thr	655	Cys	Arg	Ser	Leu
Ser	Tyr	660	Pro	Asp	Glu	Glu	Cys	665	Asn	Glu	Ala	Cys	Leu	670	Glu	Gly	Cys	Phe
Cys	Pro	675	Pro	Gly	Leu	Tyr	Met	680	Asp	Glu	Arg	Gly	Asp	685	Cys	Val	Pro	Lys
Ala	Gln	690	Cys	Pro	Cys	Tyr	Tyr	695	Asp	Gly	Glu	Ile	Phe	700	Gln	Pro	Glu	Asp
Ile	Phe	705	Ser	Asp	His	His	Thr	710	Met	Cys	Tyr	Cys	Glu	715	Asp	Gly	Phe	Met
His	Cys	725	Thr	Met	Ser	Gly	Val	730	Pro	Gly	Ser	Leu	Leu	735	Pro	Asp	Ala	Val
Leu	Ser	740	Ser	Pro	Leu	Ser	His	745	Arg	Ser	Lys	Arg	Ser	750	Leu	Ser	Cys	Arg
Pro	Pro	755	Met	Val	Lys	Leu	Val	760	Cys	Pro	Ala	Asp	Asn	765	Leu	Arg	Ala	Glu
Gly	Leu	770	Glu	Cys	Thr	Lys	Thr	775	Cys	Gln	Asn	Tyr	Asp	780	Leu	Glu	Cys	Met
Ser	Met	785	Gly	Cys	Val	Ser	Gly	790	Cys	Leu	Cys	Pro	Pro	795	Gly	Met	Val	Arg
His	Glu	805	Asn	Arg	Cys	Val	Ala	810	Leu	Glu	Arg	Cys	Pro	815	Cys	Phe	His	Gln
Gly	Lys	820	Glu	Tyr	Ala	Pro	Gly	825	Glu	Thr	Val	Lys	Ile	830	Gly	Cys	Asn	Thr
Cys	Val	835	Cys	Arg	Asp	Arg	Lys	840	Trp	Asn	Cys	Thr	Asp	845	His	Val	Cys	Asp
Ala	Thr	850	Cys	Ser	Thr	Ile	Gly	855	Met	Ala	His	Tyr	Leu	860	Thr	Phe	Asp	Gly
Leu	Lys	865	Tyr	Leu	Phe	Pro	Gly	870	Glu	Cys	Gln	Tyr	Val	875	Leu	Val	Gln	Asp
Tyr	Cys	885	Gly	Ser	Asn	Pro	Gly	890	Thr	Phe	Arg	Ile	Leu	895	Val	Gly	Asn	Lys
Gly	Cys	900	Ser	His	Pro	Ser	Val	905	Lys	Cys	Lys	Lys	Arg	910	Val	Thr	Ile	Leu
Val	Glu	915	Gly	Gly	Glu	Ile	Glu	920	Leu	Phe	Asp	Gly	Glu	925	Val	Asn	Val	Lys
Arg	Pro	930	Met	Lys	Asp	Glu	Thr	935	His	Phe	Glu	Val	Val	940	Glu	Ser	Gly	Arg
Tyr	Ile	945	Ile	Leu	Leu	Gly	Lys	950	Ala	Leu	Ser	Val	Val	955	Trp	Asp	Arg	Arg
His	Leu	965	Ser	Ile	Ser	Val	Val	970	Leu	Lys	Gln	Thr	Tyr	975	Gln	Glu	Lys	Val

CEN5301WOPCTCORRECTEDSEQLIST.txt

980 985 990
 Cys Gly Leu Cys Gly Asn Phe Asp Gly Ile Gln Asn Asn Asp Leu Thr
 995 1000 1005
 Ser Ser Asn Leu Gln Val Glu Glu Asp Pro Val Asp Phe Gly Asn Ser
 1010 1015 1020
 Trp Lys Val Ser Ser Gln Cys Ala Asp Thr Arg Lys Val Pro Leu Asp
 1025 1030 1035 1040
 Ser Ser Pro Ala Thr Cys His Asn Asn Ile Met Lys Gln Thr Met Val
 1045 1050 1055
 Asp Ser Ser Cys Arg Ile Leu Thr Ser Asp Val Phe Gln Asp Cys Asn
 1060 1065 1070
 Lys Leu Val Asp Pro Glu Pro Tyr Leu Asp Val Cys Ile Tyr Asp Thr
 1075 1080 1085
 Cys Ser Cys Glu Ser Ile Gly Asp Cys Ala Cys Phe Cys Asp Thr Ile
 1090 1095 1100
 Ala Ala Tyr Ala His Val Cys Ala Gln His Gly Lys Val Val Thr Trp
 1105 1110 1115 1120
 Arg Thr Ala Thr Leu Cys Pro Gln Ser Cys Glu Glu Arg Asn Leu Arg
 1125 1130 1135
 Glu Asn Gly Tyr Glu Cys Glu Trp Arg Tyr Asn Ser Cys Ala Pro Ala
 1140 1145 1150
 Cys Gln Val Thr Cys Gln His Pro Glu Pro Leu Ala Cys Pro Val Gln
 1155 1160 1165
 Cys Val Glu Gly Cys His Ala His Cys Pro Pro Gly Lys Ile Leu Asp
 1170 1175 1180
 Glu Leu Leu Gln Thr Cys Val Asp Pro Glu Asp Cys Pro Val Cys Glu
 1185 1190 1195 1200
 Val Ala Gly Arg Arg Phe Ala Ser Gly Lys Lys Val Thr Leu Asn Pro
 1205 1210 1215
 Ser Asp Pro Glu His Cys Gln Ile Cys His Cys Asp Val Val Asn Leu
 1220 1225 1230
 Thr Cys Glu Ala Cys Gln Glu Pro Gly Gly Leu Val Val Pro Pro Thr
 1235 1240 1245
 Asp Ala Pro Val Ser Pro Thr Thr Leu Tyr Val Glu Asp Ile Ser Glu
 1250 1255 1260
 Pro Pro Leu His Asp Phe Tyr Cys Ser Arg Leu Leu Asp Leu Val Phe
 1265 1270 1275 1280
 Leu Leu Asp Gly Ser Ser Arg Leu Ser Glu Ala Glu Phe Glu Val Leu
 1285 1290 1295
 Lys Ala Phe Val Val Asp Met Met Glu Arg Leu Arg Ile Ser Gln Lys
 1300 1305 1310
 Trp Val Arg Val Ala Val Val Glu Tyr His Asp Gly Ser His Ala Tyr
 1315 1320 1325
 Ile Gly Leu Lys Asp Arg Lys Arg Pro Ser Glu Leu Arg Arg Ile Ala
 1330 1335 1340
 Ser Gln Val Lys Tyr Ala Gly Ser Gln Val Ala Ser Thr Ser Glu Val
 1345 1350 1355 1360
 Leu Lys Tyr Thr Leu Phe Gln Ile Phe Ser Lys Ile Asp Arg Pro Glu
 1365 1370 1375
 Ala Ser Arg Ile Ala Leu Leu Leu Met Ala Ser Gln Glu Pro Gln Arg
 1380 1385 1390
 Met Ser Arg Asn Phe Val Arg Tyr Val Gln Gly Leu Lys Lys Lys
 1395 1400 1405
 Val Ile Val Ile Pro Val Gly Ile Gly Pro His Ala Asn Leu Lys Gln
 1410 1415 1420
 Ile Arg Leu Ile Glu Lys Gln Ala Pro Glu Asn Lys Ala Phe Val Leu
 1425 1430 1435 1440
 Ser Ser Val Asp Glu Leu Glu Gln Gln Arg Asp Glu Ile Val Ser Tyr
 1445 1450 1455
 Leu Cys Asp Leu Ala Pro Glu Ala Pro Pro Thr Leu Pro Pro Asp
 1460 1465 1470
 Met Ala Gln Val Thr Val Gly Pro Gly Leu Leu Gly Val Ser Thr Leu
 1475 1480 1485
 Gly Pro Lys Arg Asn Ser Met Val Leu Asp Val Ala Phe Val Leu Glu
 1490 1495 1500
 Gly Ser Asp Lys Ile Gly Glu Ala Asp Phe Asn Arg Ser Lys Glu Phe
 1505 1510 1515 1520
 Met Glu Glu Val Ile Gln Arg Met Asp Val Gly Gln Asp Ser Ile His

CEN5301WOPCTCORRECTEDSEQLIST.txt

1525 1530 1535
 Val Thr Val Leu Gln Tyr Ser Tyr Met Val Thr Val Glu Tyr Pro Phe
 1540 1545 1550
 Ser Glu Ala Gln Ser Lys Gly Asp Ile Leu Gln Arg Val Arg Glu Ile
 1555 1560 1565
 Arg Tyr Gln Gly Gly Asn Arg Thr Asn Thr Gly Leu Ala Leu Arg Tyr
 1570 1575 1580
 Leu Ser Asp His Ser Phe Leu Val Ser Gln Gly Asp Arg Glu Gln Ala
 1585 1590 1595 1600
 Pro Asn Leu Val Tyr Met Val Thr Gly Asn Pro Ala Ser Asp Glu Ile
 1605 1610 1615
 Lys Arg Leu Pro Gly Asp Ile Gln Val Val Pro Ile Gly Val Gly Pro
 1620 1625 1630
 Asn Ala Asn Val Gln Glu Leu Glu Arg Ile Gly Trp Pro Asn Ala Pro
 1635 1640 1645
 Ile Leu Ile Gln Asp Phe Glu Thr Leu Pro Arg Glu Ala Pro Asp Leu
 1650 1655 1660
 Val Leu Gln Arg Cys Cys Ser Gly Glu Gly Leu Gln Ile Pro Thr Leu
 1665 1670 1675 1680
 Ser Pro Ala Pro Asp Cys Ser Gln Pro Leu Asp Val Ile Leu Leu Leu
 1685 1690 1695
 Asp Gly Ser Ser Ser Phe Pro Ala Ser Tyr Phe Asp Glu Met Lys Ser
 1700 1705 1710
 Phe Ala Lys Ala Phe Ile Ser Lys Ala Asn Ile Gly Pro Arg Leu Thr
 1715 1720 1725
 Gln Val Ser Val Leu Gln Tyr Gly Ser Ile Thr Thr Ile Asp Val Pro
 1730 1735 1740
 Trp Asn Val Val Pro Glu Lys Ala His Leu Leu Ser Leu Val Asp Val
 1745 1750 1755 1760
 Met Gln Arg Glu Gly Gly Pro Ser Gln Ile Gly Asp Ala Leu Gly Phe
 1765 1770 1775
 Ala Val Arg Tyr Leu Thr Ser Glu Met His Gly Ala Arg Pro Gly Ala
 1780 1785 1790
 Ser Lys Ala Val Val Ile Leu Val Thr Asp Val Ser Val Asp Ser Val
 1795 1800 1805
 Asp Ala Ala Ala Asp Ala Ala Arg Ser Asn Arg Val Thr Val Phe Pro
 1810 1815 1820
 Ile Gly Ile Gly Asp Arg Tyr Asp Ala Ala Gln Leu Arg Ile Leu Ala
 1825 1830 1835 1840
 Gly Pro Ala Gly Asp Ser Asn Val Val Lys Leu Gln Arg Ile Glu Asp
 1845 1850 1855
 Leu Pro Thr Met Val Thr Leu Gly Asn Ser Phe Leu His Lys Leu Cys
 1860 1865 1870
 Ser Gly Phe Val Arg Ile Cys Met Asp Glu Asp Gly Asn Glu Lys Arg
 1875 1880 1885
 Pro Gly Asp Val Trp Thr Leu Pro Asp Gln Cys His Thr Val Thr Cys
 1890 1895 1900
 Gln Pro Asp Gly Gln Thr Leu Leu Lys Ser His Arg Val Asn Cys Asp
 1905 1910 1915 1920
 Arg Gly Leu Arg Pro Ser Cys Pro Asn Ser Gln Ser Pro Val Lys Val
 1925 1930 1935
 Glu Glu Thr Cys Gly Cys Arg Trp Thr Cys Pro Cys Val Cys Thr Gly
 1940 1945 1950
 Ser Ser Thr Arg His Ile Val Thr Phe Asp Gly Gln Asn Phe Lys Leu
 1955 1960 1965
 Thr Gly Ser Cys Ser Tyr Val Leu Phe Gln Asn Lys Glu Gln Asp Leu
 1970 1975 1980
 Glu Val Ile Leu His Asn Gly Ala Cys Ser Pro Gly Ala Arg Gln Gly
 1985 1990 1995 2000
 Cys Met Lys Ser Ile Glu Val Lys His Ser Ala Leu Ser Val Glu Leu
 2005 2010 2015
 His Ser Asp Met Glu Val Thr Val Asn Gly Arg Leu Val Ser Val Pro
 2020 2025 2030
 Tyr Val Gly Gly Asn Met Glu Val Asn Val Tyr Gly Ala Ile Met His
 2035 2040 2045
 Glu Val Arg Phe Asn His Leu Gly His Ile Phe Thr Phe Thr Pro Gln
 2050 2055 2060
 Asn Asn Glu Phe Gln Leu Gln Leu Ser Pro Lys Thr Phe Ala Ser Lys

CEN5301WOPCTCORRECTEDSEQLIST.txt

2065 2070 2075 2080
 Thr Tyr Gly Leu Cys Gly Ile Cys Asp Glu Asn Gly Ala Asn Asp Phe
 2085 2090 2095
 Met Leu Arg Asp Gly Thr Val Thr Thr Asp Trp Lys Thr Leu Val Gln
 2100 2105 2110
 Glu Trp Thr Val Gln Arg Pro Gly Gln Thr Cys Gln Pro Ile Leu Glu
 2115 2120 2125
 Glu Gln Cys Leu Val Pro Asp Ser Ser His Cys Gln Val Leu Leu Leu
 2130 2135 2140
 Pro Leu Phe Ala Glu Cys His Lys Val Leu Ala Pro Ala Thr Phe Tyr
 2145 2150 2155 2160
 Ala Ile Cys Gln Gln Asp Ser Cys His Gln Glu Gln Val Cys Glu Val
 2165 2170 2175
 Ile Ala Ser Tyr Ala His Leu Cys Arg Thr Asn Gly Val Cys Val Asp
 2180 2185 2190
 Trp Arg Thr Pro Asp Phe Cys Ala Met Ser Cys Pro Pro Ser Leu Val
 2195 2200 2205
 Tyr Asn His Cys Glu His Gly Cys Pro Arg His Cys Asp Gly Asn Val
 2210 2215 2220
 Ser Ser Cys Gly Asp His Pro Ser Glu Gly Cys Phe Cys Pro Pro Asp
 2225 2230 2235 2240
 Lys Val Met Leu Glu Gly Ser Cys Val Pro Glu Glu Ala Cys Thr Gln
 2245 2250 2255
 Cys Ile Gly Glu Asp Gly Val Gln His Gln Phe Leu Glu Ala Trp Val
 2260 2265 2270
 Pro Asp His Gln Pro Cys Gln Ile Cys Thr Cys Leu Ser Gly Arg Lys
 2275 2280 2285
 Val Asn Cys Thr Thr Gln Pro Cys Pro Thr Ala Lys Ala Pro Thr Cys
 2290 2295 2300
 Gly Leu Cys Glu Val Ala Arg Leu Arg Gln Asn Ala Asp Gln Cys Cys
 2305 2310 2315 2320
 Pro Glu Tyr Glu Cys Val Cys Asp Pro Val Ser Cys Asp Leu Pro Pro
 2325 2330 2335
 Val Pro His Cys Glu Arg Gly Leu Gln Pro Thr Leu Thr Asn Pro Gly
 2340 2345 2350
 Glu Cys Arg Pro Asn Phe Thr Cys Ala Cys Arg Lys Glu Glu Cys Lys
 2355 2360 2365
 Arg Val Ser Pro Pro Ser Cys Pro Pro His Arg Leu Pro Thr Leu Arg
 2370 2375 2380
 Lys Thr Gln Cys Cys Asp Glu Tyr Glu Cys Ala Cys Asn Cys Val Asn
 2385 2390 2395 2400
 Ser Thr Val Ser Cys Pro Leu Gly Tyr Leu Ala Ser Thr Ala Thr Asn
 2405 2410 2415
 Asp Cys Gly Cys Thr Thr Thr Cys Leu Pro Asp Lys Val Cys Val
 2420 2425 2430
 His Arg Ser Thr Ile Tyr Pro Val Gly Gln Phe Trp Glu Glu Gly Cys
 2435 2440 2445
 Asp Val Cys Thr Cys Thr Asp Met Glu Asp Ala Val Met Gly Leu Arg
 2450 2455 2460
 Val Ala Gln Cys Ser Gln Lys Pro Cys Glu Asp Ser Cys Arg Ser Gly
 2465 2470 2475 2480
 Phe Thr Tyr Val Leu His Glu Gly Glu Cys Cys Gly Arg Cys Leu Pro
 2485 2490 2495
 Ser Ala Cys Glu Val Val Thr Gly Ser Pro Arg Gly Asp Ser Gln Ser
 2500 2505 2510
 Ser Trp Lys Ser Val Gly Ser Gln Trp Ala Ser Pro Glu Asn Pro Cys
 2515 2520 2525
 Leu Ile Asn Glu Cys Val Arg Val Lys Glu Glu Val Phe Ile Gln Gln
 2530 2535 2540
 Arg Asn Val Ser Cys Pro Gln Leu Glu Val Pro Val Cys Pro Ser Gly
 2545 2550 2555 2560
 Phe Gln Leu Ser Cys Lys Thr Ser Ala Cys Cys Pro Ser Cys Arg Cys
 2565 2570 2575
 Glu Arg Met Glu Ala Cys Met Leu Asn Gly Thr Val Ile Gly Pro Gly
 2580 2585 2590
 Lys Thr Val Met Ile Asp Val Cys Thr Thr Cys Arg Cys Met Val Gln
 2595 2600 2605
 Val Gly Val Ile Ser Gly Phe Lys Leu Glu Cys Arg Lys Thr Thr Cys

CEN5301WOPCTCORRECTEDSEQLIST.txt

2610
 Asn Pro Cys Pro Leu Gly Tyr Lys Glu Glu Asn Asn Thr Gly Glu Cys
 2625 2630 2635 2640
 Cys Gly Arg Cys Leu Pro Thr Ala Cys Thr Ile Gln Leu Arg Gly Gly
 2645 2650 2655
 Gln Ile Met Thr Leu Lys Arg Asp Glu Thr Leu Gln Asp Gly Cys Asp
 2660 2665 2670
 Thr His Phe Cys Lys Val Asn Glu Arg Gly Glu Tyr Phe Trp Glu Lys
 2675 2680 2685
 Arg Val Thr Gly Cys Pro Pro Phe Asp Glu His Lys Cys Leu Ala Glu
 2690 2695 2700
 Gly Gly Lys Ile Met Lys Ile Pro Gly Thr Cys Cys Asp Thr Cys Glu
 2705 2710 2715 2720
 Glu Pro Glu Cys Asn Asp Ile Thr Ala Arg Leu Gln Tyr Val Lys Val
 2725 2730 2735
 Gly Ser Cys Lys Ser Glu Val Glu Val Asp Ile His Tyr Cys Gln Gly
 2740 2745 2750
 Lys Cys Ala Ser Lys Ala Met Tyr Ser Ile Asp Ile Asn Asp Val Gln
 2755 2760 2765
 Asp Gln Cys Ser Cys Cys Ser Pro Thr Arg Thr Glu Pro Met Gln Val
 2770 2775 2780
 Ala Leu His Cys Thr Asn Gly Ser Val Val Tyr His Glu Val Leu Asn
 2785 2790 2795 2800
 Ala Met Glu Cys Lys Cys Ser Pro Arg Lys Cys Ser Lys
 2805 2810

<210> 58
 <211> 277
 <212> PRT
 <213> Homo sapiens

<400> 58
 Met Val Arg Leu Pro Leu Gln Cys Val Leu Trp Gly Cys Leu Leu Thr
 1 5 10 15
 Ala Val His Pro Glu Pro Pro Thr Ala Cys Arg Glu Lys Gln Tyr Leu
 20 25 30
 Ile Asn Ser Gln Cys Cys Ser Leu Cys Gln Pro Gly Gln Lys Leu Val
 35 40 45
 Ser Asp Cys Thr Glu Phe Thr Glu Thr Glu Cys Leu Pro Cys Gly Glu
 50 55 60
 Ser Glu Phe Leu Asp Thr Trp Asn Arg Glu Thr His Cys His Gln His
 65 70 75 80
 Lys Tyr Cys Asp Pro Asn Leu Gly Leu Arg Val Gln Gln Lys Gly Thr
 85 90 95
 Ser Glu Thr Asp Thr Ile Cys Thr Cys Glu Glu Gly Trp His Cys Thr
 100 105 110
 Ser Glu Ala Cys Glu Ser Cys Val Leu His Arg Ser Cys Ser Pro Gly
 115 120 125
 Phe Gly Val Lys Gln Ile Ala Thr Gly Val Ser Asp Thr Ile Cys Glu
 130 135 140
 Pro Cys Pro Val Gly Phe Phe Ser Asn Val Ser Ser Ala Phe Glu Lys
 145 150 155 160
 Cys His Pro Trp Thr Ser Cys Glu Thr Lys Asp Leu Val Val Gln Gln
 165 170 175
 Ala Gly Thr Asn Lys Thr Asp Val Val Cys Gly Pro Gln Asp Arg Leu
 180 185 190
 Arg Ala Leu Val Val Ile Pro Ile Ile Phe Gly Ile Leu Phe Ala Ile
 195 200 205
 Leu Leu Val Leu Val Phe Ile Lys Lys Val Ala Lys Lys Pro Thr Asn
 210 215 220
 Lys Ala Pro His Pro Lys Gln Glu Pro Gln Glu Ile Asn Phe Pro Asp
 225 230 235 240
 Asp Leu Pro Gly Ser Asn Thr Ala Ala Pro Val Gln Glu Thr Leu His
 245 250 255
 Gly Cys Gln Pro Val Thr Gln Glu Asp Gly Lys Glu Ser Arg Ile Ser
 260 265 270
 Val Gln Glu Arg Gln

275

<210> 59
 <211> 273
 <212> PRT
 <213> Homo sapiens

<400> 59
 Met Lys Lys Thr Gln Thr Trp Ile Leu Thr Cys Ile Tyr Leu Gln Leu
 1 5 10 15
 Leu Leu Phe Asn Pro Leu Val Lys Thr Glu Gly Ile Cys Arg Asn Arg
 20 25 30
 Val Thr Asn Asn Val Lys Asp Val Thr Lys Leu Val Ala Asn Leu Pro
 35 40 45
 Lys Asp Tyr Met Ile Thr Leu Lys Tyr Val Pro Gly Met Asp Val Leu
 50 55 60
 Pro Ser His Cys Trp Ile Ser Glu Met Val Val Gln Leu Ser Asp Ser
 65 70 75 80
 Leu Thr Asp Leu Leu Asp Lys Phe Ser Asn Ile Ser Glu Gly Leu Ser
 85 90 95
 Asn Tyr Ser Ile Ile Asp Lys Leu Val Asn Ile Val Asp Asp Leu Val
 100 105 110
 Glu Cys Val Lys Glu Asn Ser Ser Lys Asp Leu Lys Lys Ser Phe Lys
 115 120 125
 Ser Pro Glu Pro Arg Leu Phe Thr Pro Glu Glu Phe Phe Arg Ile Phe
 130 135 140
 Asn Arg Ser Ile Asp Ala Phe Lys Asp Phe Val Val Ala Ser Glu Thr
 145 150 155 160
 Ser Asp Cys Val Val Ser Ser Thr Leu Ser Pro Glu Lys Asp Ser Arg
 165 170 175
 Val Ser Val Thr Lys Pro Phe Met Leu Pro Pro Val Ala Ala Ser Ser
 180 185 190
 Leu Arg Asn Asp Ser Ser Ser Ser Asn Arg Lys Ala Lys Asn Pro Pro
 195 200 205
 Gly Asp Ser Ser Leu His Trp Ala Ala Met Ala Leu Tyr Trp Lys Lys Arg
 210 215 220 240
 Ser Leu Ile Ile Gly Phe Ala Phe Gly Ala Leu Tyr Trp Lys Lys Arg
 225 230 235 240
 Gln Pro Ser Leu Thr Arg Ala Val Glu Asn Ile Gln Ile Asn Glu Glu
 245 250 255
 Asp Asn Glu Ile Ser Met Leu Gln Glu Lys Glu Arg Glu Phe Gln Glu
 260 265 270
 Val

<210> 60
 <211> 177
 <212> PRT
 <213> Homo sapiens

<400> 60
 Met Phe His Val Ser Phe Arg Tyr Ile Phe Gly Leu Pro Pro Leu Ile
 1 5 10 15
 Leu Val Leu Leu Pro Val Ala Ser Ser Asp Cys Asp Ile Glu Gly Lys
 20 25 30
 Asp Gly Lys Gln Tyr Glu Ser Val Leu Met Val Ser Ile Asp Gln Leu
 35 40 45
 Leu Asp Ser Met Lys Glu Ile Gly Ser Asn Cys Leu Asn Asn Glu Phe
 50 55 60
 Asn Phe Phe Lys Arg His Ile Cys Asp Ala Asn Lys Glu Gly Met Phe
 65 70 75 80
 Leu Phe Arg Ala Ala Arg Lys Leu Arg Gln Phe Leu Lys Met Asn Ser
 85 90 95
 Thr Gly Asp Phe Asp Leu His Leu Leu Lys Val Ser Glu Gly Thr Thr
 100 105 110
 Ile Leu Leu Asn Cys Thr Gly Gln Val Lys Gly Arg Lys Pro Ala Ala
 Page 44

CEN5301WOPCTCORRECTEDSEQLIST.txt

115 120 125
 Leu Gly Glu Ala Gln Pro Thr Lys Ser Leu Glu Glu Asn Lys Ser Leu
 130 135 140
 Lys Glu Gln Lys Lys Leu Asn Asp Leu Cys Phe Leu Lys Arg Leu Leu
 145 150 155 160
 Gln Glu Ile Lys Thr Cys Trp Asn Lys Ile Leu Met Gly Thr Lys Glu
 165 170 175
 His

<210> 61
 <211> 345
 <212> PRT
 <213> Homo sapiens

<400> 61
 Met Ile Ser Pro Val Leu Ile Leu Phe Ser Ser Phe Leu Cys His Val
 1 5 10 15
 Ala Ile Ala Gly Arg Thr Cys Pro Lys Pro Asp Asp Leu Pro Phe Ser
 20 25 30
 Thr Val Val Pro Leu Lys Thr Phe Tyr Glu Pro Gly Glu Glu Ile Thr
 35 40 45
 Tyr Ser Cys Lys Pro Gly Tyr Val Ser Arg Gly Gly Met Arg Lys Phe
 50 55 60
 Ile Cys Pro Leu Thr Gly Leu Trp Pro Ile Asn Thr Leu Lys Cys Thr
 65 70 75 80
 Pro Arg Val Cys Pro Phe Ala Gly Ile Leu Glu Asn Gly Ala Val Arg
 85 90 95
 Tyr Thr Thr Phe Glu Tyr Pro Asn Thr Ile Ser Phe Ser Cys Asn Thr
 100 105 110
 Gly Phe Tyr Leu Asn Gly Ala Asp Ser Ala Lys Cys Thr Glu Glu Gly
 115 120 125
 Lys Trp Ser Pro Glu Leu Pro Val Cys Ala Pro Ile Ile Cys Pro Pro
 130 135 140
 Pro Ser Ile Pro Thr Phe Ala Thr Leu Arg Val Tyr Lys Pro Ser Ala
 145 150 155 160
 Gly Asn Asn Ser Leu Tyr Arg Asp Thr Ala Val Phe Glu Cys Leu Pro
 165 170 175
 Gln His Ala Met Phe Gly Asn Asp Thr Ile Thr Cys Thr Thr His Gly
 180 185 190
 Asn Trp Thr Lys Leu Pro Glu Cys Arg Glu Val Lys Cys Pro Phe Pro
 195 200 205
 Ser Arg Pro Asp Asn Gly Phe Val Asn Tyr Pro Ala Lys Pro Thr Leu
 210 215 220
 Tyr Tyr Lys Asp Lys Ala Thr Phe Gly Cys His Asp Gly Tyr Ser Leu
 225 230 235 240
 Asp Gly Pro Glu Glu Ile Glu Cys Thr Lys Leu Gly Asn Trp Ser Ala
 245 250 255
 Met Pro Ser Cys Lys Ala Ser Cys Lys Val Pro Val Lys Lys Ala Thr
 260 265 270
 Val Val Tyr Gln Gly Glu Arg Val Lys Ile Gln Glu Lys Phe Lys Asn
 275 280 285
 Gly Met Leu His Gly Asp Lys Val Ser Phe Phe Cys Lys Asn Lys Glu
 290 295 300
 Lys Lys Cys Ser Tyr Thr Glu Asp Ala Gln Cys Ile Asp Gly Thr Ile
 305 310 315 320
 Glu Val Pro Lys Cys Phe Lys Glu His Ser Ser Leu Ala Phe Trp Lys
 325 330 335
 Thr Asp Ala Ser Asp Val Lys Pro Cys
 340 345

<210> 62
 <211> 120
 <212> PRT
 <213> Homo sapiens

CEN5301WOPCTCORRECTEDSEQLIST.txt

<400> 62

```

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

```

<210> 63

<211> 92

<212> PRT

<213> Homo sapiens

<400> 63

```

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

```

<210> 64

<211> 404

<212> PRT

<213> Homo sapiens

<400> 64

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

180      185      190
Met Val Thr Pro Ala Arg Gly Gly Asp Pro Arg Pro Thr Phe Ser Cys
195      200      205
Ser Phe Ser Pro Gly Leu Pro Arg His Arg Ala Leu Arg Thr Ala Pro
210      215      220
Ile Gln Pro Arg Val Trp Glu Pro Val Pro Leu Glu Glu Val Gln Leu
225      230      235
Val Val Glu Pro Glu Gly Gly Ala Val Ala Pro Gly Gly Thr Val Thr
245      250      255
Leu Thr Cys Glu Val Pro Ala Gln Pro Ser Pro Gln Ile His Trp Met
260      265      270
Lys Asp Gly Val Pro Leu Pro Leu Pro Pro Ser Pro Val Leu Ile Leu
275      280      285
Pro Glu Ile Gly Pro Gln Asp Gln Gly Thr Tyr Ser Cys Val Ala Thr
290      295      300
His Ser Ser His Gly Pro Gln Glu Ser Arg Ala Val Ser Ile Ser Ile
305      310      315
Ile Glu Pro Gly Glu Glu Gly Pro Thr Ala Gly Ser Val Gly Gly Ser
325      330      335
Gly Leu Gly Thr Leu Ala Leu Ala Leu Gly Ile Leu Gly Gly Leu Gly
340      345      350
Thr Ala Ala Leu Leu Ile Gly Val Ile Leu Trp Gln Arg Arg Gln Arg
355      360      365
Arg Gly Glu Glu Arg Lys Ala Pro Glu Asn Gln Glu Glu Glu Glu
370      375      380
Arg Ala Glu Leu Asn Gln Ser Glu Glu Pro Glu Ala Gly Glu Ser Ser
385      390      395      400
Thr Gly Gly Pro

```

<210> 65

<211> 575

<212> PRT

<213> Homo sapiens

<400> 65

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 66
 <211> 99
 <212> PRT
 <213> Homo sapiens

```

<400> 66
Met Thr Ser Lys Leu Ala Val Ala Leu Leu Ala Ala Phe Leu Ile Ser
  1      5      10      15
Ala Ala Leu Cys Glu Gly Ala Val Leu Pro Arg Ser Ala Lys Glu Leu
  20      25      30
Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro Phe His Pro Lys Phe
  35      40      45
Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro His Cys Ala Asn Thr
  50      55      60
Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu Leu Cys Leu Asp Pro
  65      70      75      80
Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys Phe Leu Lys Arg Ala
  85      90      95
Glu Asn Ser

```

<210> 67
 <211> 100
 <212> PRT

<213> Homo sapiens

<400> 67

```

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

```

<210> 68

<211> 1474

<212> PRT

<213> Homo sapiens

<400> 68

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

Val	Lys	Val	340	Asp	Ser	His	Phe	Arg	345	Gln	Gly	Ile	Pro	Phe	350	Phe	Gly	Gln
Val	Arg	Leu	355	Val	Asp	Gly	Lys	Gly	360	Val	Pro	Ile	Pro	Asn	365	Lys	Val	Ile
Phe	Ile	Arg	370	Gly	Asn	Glu	Ala	Asn	375	Tyr	Tyr	Ser	Asn	Ala	380	Thr	Thr	Asp
Glu	His	Gly	385	Leu	Val	Gln	Phe	Ser	390	Ile	Asn	Thr	Thr	Asn	400	Val	Met	Gly
Thr	Ser	Leu	405	Thr	Val	Arg	Val	Asn	410	Tyr	Lys	Asp	Arg	Ser	415	Pro	Cys	Tyr
Gly	Tyr	Gln	420	Trp	Val	Ser	Glu	Glu	425	His	Glu	Glu	Ala	His	430	His	Thr	Ala
Tyr	Leu	Val	435	Phe	Ser	Pro	Ser	Lys	440	Ser	Phe	Val	His	Leu	445	Glu	Pro	Met
Ser	His	Glu	450	Leu	Pro	Cys	Gly	His	455	Thr	Gln	Thr	Val	Gln	460	Ala	His	Tyr
Ile	Leu	Asn	465	Gly	Gly	Thr	Leu	Leu	470	Gly	Leu	Lys	Lys	Leu	475	Ser	Phe	Tyr
Tyr	Leu	Ile	485	Met	Ala	Lys	Gly	Gly	490	Ile	Val	Arg	Thr	Gly	495	Thr	His	Gly
Leu	Leu	Val	500	Lys	Gln	Glu	Asp	Met	505	Lys	Gly	His	Phe	Ser	510	Ile	Ser	Ile
Pro	Val	Lys	515	Ser	Asp	Ile	Ala	Pro	520	Val	Ala	Arg	Leu	Leu	525	Ile	Tyr	Ala
Val	Leu	Pro	530	Thr	Gly	Asp	Val	Ile	535	Gly	Asp	Ser	Ala	Lys	540	Tyr	Asp	Val
Glu	Asn	Cys	545	Leu	Ala	Asn	Lys	Val	550	Asp	Leu	Ser	Phe	Ser	555	Pro	Ser	Gln
Ser	Leu	Pro	565	Ala	Ser	His	Ala	His	570	Leu	Arg	Val	Thr	Ala	575	Ala	Pro	Gln
Ser	Val	Cys	580	Ala	Leu	Arg	Ala	Val	585	Asp	Gln	Ser	Val	Leu	590	Leu	Met	Lys
Pro	Asp	Ala	595	Glu	Leu	Ser	Ala	Ser	600	Ser	Val	Tyr	Asn	Leu	605	Leu	Pro	Glu
Lys	Asp	Leu	610	Thr	Gly	Phe	Pro	Gly	615	Pro	Leu	Asn	Asp	Gln	620	Asp	Asp	Glu
Asp	Cys	Ile	625	Asn	Arg	His	Asn	Val	630	Tyr	Ile	Asn	Gly	Ile	635	Thr	Tyr	Thr
Pro	Val	Ser	645	Ser	Thr	Asn	Glu	Lys	650	Asp	Met	Tyr	Ser	Phe	655	Leu	Glu	Asp
Met	Gly	Leu	660	Lys	Ala	Phe	Thr	Asn	665	Ser	Lys	Ile	Arg	Lys	670	Pro	Lys	Met
Cys	Pro	Gln	675	Leu	Gln	Gln	Tyr	Glu	680	Met	His	Gly	Pro	Glu	685	Gly	Leu	Arg
Val	Gly	Phe	690	Tyr	Glu	Ser	Asp	Val	695	Met	Gly	Arg	Gly	His	700	Ala	Arg	Leu
Val	His	Val	705	Glu	Glu	Pro	His	Thr	710	Glu	Thr	Val	Arg	Lys	715	Tyr	Phe	Pro
Glu	Thr	Trp	725	Ile	Trp	Asp	Leu	Val	730	Val	Val	Asn	Ser	Ala	735	Gly	Val	Ala
Glu	Val	Gly	740	Val	Thr	Val	Pro	Asp	745	Thr	Ile	Thr	Glu	Trp	750	Lys	Ala	Gly
Ala	Phe	Cys	755	Leu	Ser	Glu	Asp	Ala	760	Gly	Leu	Gly	Ile	Ser	765	Ser	Thr	Ala
Ser	Leu	Arg	770	Ala	Phe	Gln	Pro	Phe	775	Phe	Val	Glu	Leu	Thr	780	Met	Pro	Tyr
Ser	Val	Ile	785	Arg	Gly	Glu	Ala	Phe	790	Thr	Leu	Lys	Ala	Thr	795	Val	Leu	Asn
Tyr	Leu	Pro	805	Lys	Cys	Ile	Arg	Val	810	Ser	Val	Gln	Leu	Glu	815	Ala	Ser	Pro
Ala	Phe	Leu	820	Ala	Val	Pro	Val	Glu	825	Lys	Glu	Gln	Ala	Pro	830	His	Cys	Ile
Cys	Ala	Asn	835	Gly	Arg	Gln	Thr	Val	840	Ser	Trp	Ala	Val	Thr	845	Pro	Lys	Ser
Leu	Gly	Asn	850	Val	Asn	Phe	Thr	Val	855	Ser	Ala	Glu	Ala	Leu	860	Glu	Ser	Gln
Glu	Leu	Cys	865	Gly	Thr	Glu	Val	Pro	870	Ser	Val	Pro	Glu	His	875	Gly	Arg	Lys

CEN5301WOPCTCORRECTEDSEQLIST.txt

885 890 895
 Asp Thr Val Ile Lys Pro Leu Leu Val Glu Pro Glu Gly Leu Glu Lys
 900 905 910
 Glu Thr Thr Phe Asn Ser Leu Leu Cys Pro Ser Gly Gly Glu Val Ser
 915 920 925
 Glu Glu Leu Ser Leu Lys Leu Pro Pro Asn Val Val Glu Glu Ser Ala
 930 935 940
 Arg Ala Ser Val Ser Val Leu Gly Asp Ile Leu Gly Ser Ala Met Gln
 945 950 955 960
 Asn Thr Gln Asn Leu Leu Gln Met Pro Tyr Gly Cys Gly Glu Gln Asn
 965 970 975
 Met Val Leu Phe Ala Pro Asn Ile Tyr Val Leu Asp Tyr Leu Asn Glu
 980 985 990
 Thr Gln Gln Leu Thr Pro Glu Ile Lys Ser Lys Ala Ile Gly Tyr Leu
 995 1000 1005
 Asn Thr Gly Tyr Gln Arg Gln Leu Asn Tyr Lys His Tyr Asp Gly Ser
 1010 1015 1020
 Tyr Ser Thr Phe Gly Glu Arg Tyr Gly Arg Asn Gln Gly Asn Thr Trp
 1025 1030 1035 1040
 Leu Thr Ala Phe Val Leu Lys Thr Phe Ala Gln Ala Arg Ala Tyr Ile
 1045 1050 1055
 Phe Ile Asp Glu Ala His Ile Thr Gln Ala Leu Ile Trp Leu Ser Gln
 1060 1065 1070
 Arg Gln Lys Asp Asn Gly Cys Phe Arg Ser Ser Gly Ser Leu Leu Asn
 1075 1080 1085
 Asn Ala Ile Lys Gly Gly Val Glu Asp Glu Val Thr Leu Ser Ala Tyr
 1090 1095 1100
 Ile Thr Ile Ala Leu Leu Glu Ile Pro Leu Thr Val Thr His Pro Val
 1105 1110 1115 1120
 Val Arg Asn Ala Leu Phe Cys Leu Glu Ser Ala Trp Lys Thr Ala Gln
 1125 1130 1135
 Glu Gly Asp His Gly Ser His Val Tyr Thr Lys Ala Leu Leu Ala Tyr
 1140 1145 1150
 Ala Phe Ala Leu Ala Gly Asn Gln Asp Lys Arg Lys Glu Val Leu Lys
 1155 1160 1165
 Ser Leu Asn Glu Glu Ala Val Lys Lys Asp Asn Ser Val His Trp Glu
 1170 1175 1180
 Arg Pro Gln Lys Pro Lys Ala Pro Val Gly His Phe Tyr Glu Pro Gln
 1185 1190 1195 1200
 Ala Pro Ser Ala Glu Val Glu Met Thr Ser Tyr Val Leu Leu Ala Tyr
 1205 1210 1215
 Leu Thr Ala Gln Pro Ala Pro Thr Ser Glu Asp Leu Thr Ser Ala Thr
 1220 1225 1230
 Asn Ile Val Lys Trp Ile Thr Lys Gln Gln Asn Ala Gln Gly Gly Phe
 1235 1240 1245
 Ser Ser Thr Gln Asp Thr Val Val Ala Leu His Ala Leu Ser Lys Tyr
 1250 1255 1260
 Gly Ala Ala Thr Phe Thr Arg Thr Gly Lys Ala Ala Gln Val Thr Ile
 1265 1270 1275 1280
 Gln Ser Ser Gly Thr Phe Ser Ser Lys Phe Gln Val Asp Asn Asn Asn
 1285 1290 1295
 Arg Leu Leu Leu Gln Gln Val Ser Leu Pro Glu Leu Pro Gly Glu Tyr
 1300 1305 1310
 Ser Met Lys Val Thr Gly Glu Gly Cys Val Tyr Leu Gln Thr Ser Leu
 1315 1320 1325
 Lys Tyr Asn Ile Leu Pro Glu Lys Glu Glu Phe Pro Phe Ala Leu Gly
 1330 1335 1340
 Val Gln Thr Leu Pro Gln Thr Cys Asp Glu Pro Lys Ala His Thr Ser
 1345 1350 1355 1360
 Phe Gln Ile Ser Leu Ser Val Ser Tyr Thr Gly Ser Arg Ser Ala Ser
 1365 1370 1375
 Asn Met Ala Ile Val Asp Val Lys Met Val Ser Gly Phe Ile Pro Leu
 1380 1385 1390
 Lys Pro Thr Val Lys Met Leu Glu Arg Ser Asn His Val Ser Arg Thr
 1395 1400 1405
 Glu Val Ser Ser Asn His Val Leu Ile Tyr Leu Asp Lys Val Ser Asn
 1410 1415 1420
 Gln Thr Leu Ser Leu Phe Phe Thr Val Leu Gln Asp Val Pro Val Arg

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 71
 <211> 412
 <212> PRT
 <213> Homo sapiens

```

<400> 71
Met Ala Pro Pro Ser Val Phe Ala Glu Val Pro Gln Ala Gln Pro Val
1      5      10      15
Leu Val Phe Lys Leu Thr Ala Asp Phe Arg Glu Asp Pro Asp Pro Arg
20      25      30
Lys Val Asn Leu Gly Val Gly Ala Tyr Arg Thr Asp Asp Cys His Pro
35      40      45
Trp Val Leu Pro Val Val Lys Lys Val Glu Gln Lys Ile Ala Asn Asp
50      55      60
Asn Ser Leu Asn His Glu Tyr Leu Pro Ile Leu Gly Leu Ala Glu Phe
65      70      75      80
Arg Ser Cys Ala Ser Arg Leu Ala Leu Gly Asp Asp Ser Pro Ala Leu
85      90      95
Lys Glu Lys Arg Val Gly Gly Val Gln Ser Leu Gly Gly Thr Gly Ala
100      105      110
Leu Arg Ile Gly Ala Asp Phe Leu Ala Arg Trp Tyr Asn Gly Thr Asn
115      120      125
Asn Lys Asn Thr Pro Val Tyr Val Ser Ser Pro Thr Trp Glu Asn His
130      135      140
Asn Ala Val Phe Ser Ala Ala Gly Phe Lys Asp Ile Arg Ser Tyr Arg
145      150      155      160
Tyr Trp Asp Ala Glu Lys Arg Gly Leu Asp Leu Gln Gly Phe Leu Asn
165      170      175      180
Asp Leu Glu Asn Ala Pro Glu Phe Ser Ile Val Val Leu His Ala Cys
185      190      195
Ala His Asn Pro Thr Gly Ile Asp Pro Thr Pro Glu Gln Trp Lys Gln
200      205      210
Ile Ala Ser Val Met Lys His Arg Phe Leu Phe Pro Phe Phe Asp Ser
215      220      225

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

210	Ala	Tyr	Gln	Gly	Phe	Ala	215	Ser	Gly	Asn	Leu	Glu	220	Arg	Asp	Ala	Trp	Ala
225	Ile	Arg	Tyr	Phe	Val	230	Ser	Glu	Gly	Phe	Glu	235	Phe	Phe	Cys	Ala	Gln	Ser
				245							250	Glu	Arg	Val	Gly	Asn	Leu	Thr
	Phe	Ser	Lys	Asn	Phe	Gly	Leu	Tyr	Asn	265	Glu	Arg	Val	Gly	Asn	Leu	Thr	
	Val	Val	Gly	Lys	Glu	Pro	Glu	Ser	Ile	280	Leu	Gln	Val	Leu	Ser	Gln	Met	
	Glu	Lys	Ile	Val	Arg	Ile	Thr	Trp	Ser	Asn	Pro	Pro	Ala	Gln	Gly	Ala		
	Arg	Ile	Val	Ala	Ser	Thr	Leu	Ser	Asn	Pro	Glu	Leu	Phe	Glu	Glu	Trp		
305	Thr	Gly	Asn	Val	Lys	Thr	Met	Ala	Asp	Arg	Ile	Leu	Thr	Met	Arg	Ser		
				325						330								
	Glu	Leu	Arg	Ala	Arg	Leu	Glu	Ala	Leu	Lys	Thr	Pro	Gly	Thr	Trp	Asn		
				340						345								
	His	Ile	Thr	Asp	Gln	Ile	Gly	Met	Phe	Ser	Phe	Thr	Gly	Leu	Asn	Pro		
				355						360								
	Lys	Gln	Val	Glu	Tyr	Leu	Val	Asn	Glu	Lys	His	Ile	Tyr	Leu	Leu	Pro		
				370						375								
	Ser	Gly	Arg	Ile	Asn	Val	Ser	Gly	Leu	Thr	Thr	Lys	Asn	Leu	Asp	Tyr		
385						390												
	Val	Ala	Thr	Ser	Ile	His	Glu	Ala	Val	Thr	Lys	Ile						
					405					410								

<210> 72
 <211> 199
 <212> PRT
 <213> Homo sapiens

<400> 72	Met	Asn	Cys	Val	Cys	Arg	Leu	Val	Leu	Val	Val	Leu	Ser	Leu	Trp	Pro
1	Asp	Thr	Ala	Val	Ala	Pro	Gly	Pro	Pro	Pro	Gly	Pro	Pro	Arg	Val	Ser
			20						25					30		
	Pro	Asp	Pro	Arg	Ala	Glu	Leu	Asp	Thr	Val	Leu	Leu	Thr	Arg	Ser	
			35					40				45				
	Leu	Leu	Ala	Asp	Thr	Arg	Gln	Leu	Ala	Ala	Gln	Leu	Arg	Asp	Lys	Phe
			50				55				60					
	Pro	Ala	Asp	Gly	Asp	His	Asn	Leu	Asp	Ser	Leu	Pro	Thr	Leu	Ala	Met
65					70					75						80
	Ser	Ala	Gly	Ala	Leu	Gly	Ala	Leu	Gln	Leu	Pro	Gly	Val	Leu	Thr	Arg
				85					90					95		
	Leu	Arg	Ala	Asp	Leu	Leu	Ser	Tyr	Leu	Arg	His	Val	Gln	Trp	Leu	Arg
			100						105				110			
	Arg	Ala	Gly	Gly	Ser	Ser	Leu	Lys	Thr	Leu	Glu	Pro	Glu	Leu	Gly	Thr
			115					120					125			
	Leu	Gln	Ala	Arg	Leu	Asp	Arg	Leu	Leu	Arg	Arg	Leu	Gln	Leu	Leu	Met
			130				135					140				
	Ser	Arg	Leu	Ala	Leu	Pro	Gln	Pro	Pro	Pro	Asp	Pro	Pro	Ala	Pro	Pro
145					150					155						160
	Leu	Ala	Pro	Pro	Ser	Ala	Trp	Gly	Gly	Ile	Arg	Ala	Ala	His	Ala	
				165					170					175		
	Ile	Leu	Gly	Gly	Leu	His	Leu	Thr	Leu	Asp	Trp	Ala	Val	Arg	Gly	Leu
			180						185					190		
	Leu	Leu	Leu	Lys	Thr	Arg	Leu									
			195													

<210> 73
 <211> 153
 <212> PRT
 <213> Homo sapiens

<400> 73
 Met Gly Lys Ile Ser Ser Leu Pro Thr Gln Leu Phe Lys Cys Cys Phe

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```

<210> 74
 <211> 1304
 <212> PRT
 <213> Homo sapiens

```

<400> 74
Met Gln Leu Lys Ile Met Pro Lys Lys Lys Arg Leu Ser Ala Gly Arg
1      5      10      15
Val Pro Leu Ile Leu Phe Leu Cys Gln Met Ile Ser Ala Leu Glu Val
20      25      30
Pro Leu Asp Pro Lys Leu Leu Glu Asp Leu Val Gln Pro Pro Thr Ile
35      40      45
Thr Gln Gln Ser Pro Lys Asp Tyr Ile Ile Asp Pro Arg Glu Asn Ile
50      55      60
Val Ile Gln Cys Glu Ala Lys Gly Lys Pro Pro Pro Ser Phe Ser Trp
65      70      75      80
Thr Arg Asn Gly Thr His Phe Asp Ile Asp Lys Asp Pro Leu Val Thr
85      90      95
Met Lys Pro Gly Thr Gly Thr Leu Ile Ile Asn Ile Met Ser Glu Gly
100      105      110
Lys Ala Glu Thr Tyr Glu Gly Val Tyr Gln Cys Thr Ala Arg Asn Glu
115      120      125
Arg Gly Ala Ala Val Ser Asn Asn Ile Val Val Arg Pro Ser Arg Ser
130      135      140
Pro Leu Trp Thr Lys Glu Lys Leu Glu Pro Ile Thr Leu Gln Ser Gly
145      150      155      160
Gln Ser Leu Val Leu Pro Cys Arg Pro Pro Ile Gly Leu Pro Pro Pro
165      170      175
Ile Ile Phe Trp Met Asp Asn Ser Phe Gln Arg Leu Pro Gln Ser Glu
180      185      190
Arg Val Ser Gln Gly Leu Asn Gly Asp Leu Tyr Phe Ser Asn Val Leu
195      200      205
Pro Glu Asp Thr Arg Glu Asp Tyr Ile Cys Tyr Ala Arg Phe Asn His
210      215      220
Thr Gln Thr Ile Gln Gln Lys Gln Pro Ile Ser Val Lys Val Ile Ser
225      230      235      240
Val Asp Glu Leu Asn Asp Thr Ile Ala Ala Asn Leu Ser Asp Thr Glu
245      250      255
Phe Tyr Gly Ala Lys Ser Ser Arg Glu Arg Pro Pro Thr Phe Leu Thr
260      265      270
Pro Glu Gly Asn Ala Ser Asn Lys Glu Glu Leu Arg Gly Asn Val Leu
275      280      285
Ser Leu Glu Cys Ile Ala Glu Gly Leu Pro Thr Pro Ile Ile Tyr Trp
290      295      300
Ala Lys Glu Asp Gly Met Leu Pro Lys Asn Arg Thr Val Tyr Lys Asn
305      310      315      320
Phe Glu Lys Thr Leu Gln Ile Ile His Val Ser Glu Ala Asp Ser Gly

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

325      330      335
Asn Tyr Gln Cys Ile Ala Lys Asn Ala Leu Gly Ala Ile His His Thr
340      345      350
Ile Ser Val Arg Val Lys Ala Ala Pro Tyr Trp Ile Thr Ala Pro Gln
355      360      365
Asn Leu Val Leu Ser Pro Gly Glu Asp Gly Thr Leu Ile Cys Arg Ala
370      375      380
Asn Gly Asn Pro Lys Pro Arg Ile Ser Trp Leu Thr Asn Gly Val Pro
385      390      400
Ile Glu Ile Ala Pro Asp Asp Pro Ser Arg Lys Ile Asp Gly Asp Thr
405      410      415
Ile Ile Phe Ser Asn Val Gln Glu Arg Ser Ser Ala Val Tyr Gln Cys
420      425      430
Asn Ala Ser Asn Glu Tyr Gly Tyr Leu Leu Ala Asn Ala Phe Val Asn
435      440      445
Val Leu Ala Glu Pro Pro Arg Ile Leu Thr Pro Ala Asn Thr Leu Tyr
450      455      460
Gln Val Ile Ala Asn Arg Pro Ala Leu Leu Asp Cys Ala Phe Phe Gly
465      470      475
Ser Pro Leu Pro Thr Ile Glu Trp Phe Lys Gly Ala Lys Gly Ser Ala
485      490      495
Leu His Glu Asp Ile Tyr Val Leu His Glu Asn Gly Thr Leu Glu Ile
500      505      510
Pro Val Ala Gln Lys Asp Ser Thr Gly Thr Tyr Thr Cys Val Ala Arg
515      520      525
Asn Lys Leu Gly Met Ala Lys Asn Glu Val His Leu Glu Ile Lys Asp
530      535      540
Pro Thr Trp Ile Val Lys Gln Pro Glu Tyr Ala Val Val Gln Arg Gly
545      550      555
Ser Met Val Ser Phe Glu Cys Lys Val Lys His Asp His Thr Leu Ser
565      570      575
Leu Thr Val Leu Trp Leu Lys Asp Asn Arg Glu Leu Pro Ser Asp Glu
580      585      590
Arg Phe Thr Val Asp Lys Asp His Leu Val Val Ala Asp Val Ser Asp
595      600      605
Asp Asp Ser Gly Thr Tyr Thr Cys Val Ala Asn Thr Thr Leu Asp Ser
610      615      620
Val Ser Ala Ser Ala Val Leu Ser Val Val Ala Pro Thr Pro Thr Pro
625      630      635
Ala Pro Val Tyr Asp Val Pro Asn Pro Pro Phe Asp Leu Glu Leu Thr
645      650      655
Asp Gln Leu Asp Lys Ser Val Gln Leu Ser Trp Thr Pro Gly Asp Asp
660      665      670
Asn Asn Ser Pro Ile Thr Lys Phe Ile Ile Glu Tyr Glu Asp Ala Met
675      680      685
His Lys Pro Gly Leu Trp His His Gln Thr Glu Val Ser Gly Thr Gln
690      695      700
Thr Thr Ala Gln Leu Lys Leu Ser Pro Tyr Val Asn Tyr Ser Phe Arg
705      710      715
Val Met Ala Val Asn Ser Ile Gly Lys Ser Leu Pro Ser Glu Ala Ser
725      730      735
Glu Gln Tyr Leu Thr Lys Ala Ser Glu Pro Asp Lys Asn Pro Thr Ala
740      745      750
Val Glu Gly Leu Gly Ser Glu Pro Asp Asn Leu Val Ile Thr Trp Lys
755      760      765
Pro Leu Asn Gly Phe Glu Ser Asn Gly Pro Gly Leu Gln Tyr Lys Val
770      775      780
Ser Trp Arg Gln Lys Asp Gly Asp Asp Glu Trp Thr Ser Val Val Val
785      790      795
Ala Asn Val Ser Lys Tyr Ile Val Ser Gly Thr Pro Thr Phe Val Pro
805      810      815
Tyr Leu Ile Lys Val Gln Ala Leu Asn Asp Met Gly Phe Ala Pro Glu
820      825      830
Pro Ala Val Val Met Gly His Ser Gly Glu Asp Leu Pro Met Val Ala
835      840      845
Pro Gly Asn Val Arg Val Asn Val Val Asn Ser Thr Leu Ala Glu Val
850      855      860
His Trp Asp Pro Val Pro Leu Lys Ser Ile Arg Gly His Leu Gln Gly

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

865      870      875      880
Tyr Arg Ile Tyr Tyr Trp Lys Thr Gln Ser Ser Ser Lys Arg Asn Arg
      885      890      895
Arg His Ile Glu Lys Lys Ile Leu Thr Phe Gln Gly Ser Lys Thr His
      900      905      910
Gly Met Leu Pro Gly Leu Glu Pro Phe Ser His Tyr Thr Leu Asn Val
      915      920      925
Arg Val Val Asn Gly Lys Gly Glu Gly Pro Ala Ser Pro Asp Arg Val
      930      935      940
Phe Asn Thr Pro Glu Gly Val Pro Ser Ala Pro Ser Ser Leu Lys Ile
      945      950      955      960
Val Asn Pro Thr Leu Asp Ser Leu Thr Leu Glu Trp Asp Pro Pro Ser
      965      970      975
His Pro Asn Gly Ile Leu Thr Glu Tyr Thr Leu Lys Tyr Gln Pro Ile
      980      985      990
Asn Ser Thr His Glu Leu Gly Pro Leu Val Asp Leu Lys Ile Pro Ala
      995      1000      1005
Asn Lys Thr Arg Trp Thr Leu Lys Asn Leu Asn Phe Ser Thr Arg Tyr
      1010      1015      1020
Lys Phe Tyr Phe Tyr Ala Gln Thr Ser Ala Gly Ser Gly Ser Gln Ile
      1025      1030      1035      1040
Thr Glu Glu Ala Val Thr Thr Val Asp Glu Ala Gly Ile Leu Pro Pro
      1045      1050      1055
Asp Val Gly Ala Gly Lys Val Gln Ala Val Asn Pro Arg Ile Ser Asn
      1060      1065      1070
Leu Thr Ala Ala Ala Glu Thr Tyr Ala Asn Ile Ser Trp Glu Tyr
      1075      1080      1085
Glu Gly Pro Glu His Val Asn Phe Tyr Val Glu Tyr Gly Val Ala Gly
      1090      1095      1100
Ser Lys Glu Glu Trp Arg Lys Glu Ile Val Asn Gly Ser Arg Ser Phe
      1105      1110      1115      1120
Phe Gly Leu Lys Gly Leu Met Pro Gly Thr Ala Tyr Lys Val Arg Val
      1125      1130      1135
Gly Ala Val Gly Asp Ser Gly Phe Val Ser Ser Glu Asp Val Phe Glu
      1140      1145      1150
Thr Gly Pro Ala Met Ala Ser Arg Gln Val Asp Ile Ala Thr Gln Gly
      1155      1160      1165
Trp Phe Ile Gly Leu Met Cys Ala Val Ala Leu Leu Ile Leu Ile Leu
      1170      1175      1180
Leu Ile Val Cys Phe Ile Arg Arg Asn Lys Gly Gly Lys Tyr Pro Val
      1185      1190      1195      1200
Lys Glu Lys Glu Asp Ala His Ala Asp Pro Glu Ile Gln Pro Met Lys
      1205      1210      1215
Glu Asp Asp Gly Thr Phe Gly Glu Tyr Ser Asp Ala Glu Asp His Lys
      1220      1225      1230
Pro Leu Lys Lys Gly Ser Arg Thr Pro Ser Asp Arg Thr Val Lys Lys
      1235      1240      1245
Glu Asp Ser Asp Asp Ser Leu Val Asp Tyr Gly Glu Gly Val Asn Gly
      1250      1255      1260
Gln Phe Asn Glu Asp Gly Ser Phe Ile Gly Gln Tyr Ser Gly Lys Lys
      1265      1270      1275      1280
Glu Lys Glu Pro Ala Glu Gly Asn Glu Ser Ser Glu Ala Pro Ser Pro
      1285      1290      1295
Val Asn Ala Met Asn Ser Phe Val
      1300

```

<210> 75
 <211> 68
 <212> PRT
 <213> Homo sapiens

<400> 75
 Glu Leu Asp Val Cys Val Glu Glu Ala Glu Gly Glu Ala Pro Trp Thr
 1 5 10 15
 Trp Thr Gly Leu Cys Ile Phe Ala Ala Leu Phe Leu Leu Ser Val Ser
 20 25 30
 Tyr Ser Ala Ala Leu Thr Leu Leu Met Val Gln Arg Phe Leu Ser Ala

CEN5301WOPCTCORRECTEDSEQLIST.txt

35
 Thr Arg Gln Gly Arg Pro Gln Thr Ser Leu Asp Tyr Thr Asn Val Leu
 50
 Gln Pro His Ala
 65

<210> 76
 <211> 485
 <212> PRT
 <213> Homo sapiens

<400> 76
 Met Arg Lys Arg Ala Pro Gln Ser Glu Met Ala Pro Ala Gly Val Ser
 1 5 10 15
 Leu Arg Ala Thr Ile Leu Cys Leu Leu Ala Trp Ala Gly Leu Ala Ala
 20 25 30
 Gly Asp Arg Val Tyr Ile His Pro Phe His Leu Val Ile His Asn Glu
 35 40 45
 Ser Thr Cys Glu Gln Leu Ala Lys Ala Asn Ala Gly Lys Pro Lys Asp
 50 55 60
 Pro Thr Phe Ile Pro Ala Pro Ile Gln Ala Lys Thr Ser Pro Val Asp
 65 70 75 80
 Glu Lys Ala Leu Gln Asp Gln Leu Val Leu Val Ala Ala Lys Leu Asp
 85 90 95
 Thr Glu Asp Lys Leu Arg Ala Ala Met Val Gly Met Leu Ala Asn Phe
 100 105 110
 Leu Gly Phe Arg Ile Tyr Gly Met His Ser Glu Leu Trp Gly Val Val
 115 120 125
 His Gly Ala Thr Val Leu Ser Pro Thr Ala Val Phe Gly Thr Leu Ala
 130 135 140
 Ser Leu Tyr Leu Gly Ala Leu Asp His Thr Ala Asp Arg Leu Gln Ala
 145 150 155 160
 Ile Leu Gly Val Pro Trp Lys Asp Lys Asn Cys Thr Ser Arg Leu Asp
 165 170 175
 Ala His Lys Val Leu Ser Ala Leu Gln Ala Val Gln Gly Leu Leu Val
 180 185 190
 Ala Gln Gly Arg Ala Asp Ser Gln Ala Gln Leu Leu Leu Ser Thr Val
 195 200 205
 Val Gly Val Phe Thr Ala Pro Gly Leu His Leu Lys Gln Pro Phe Val
 210 215 220
 Gln Gly Leu Ala Leu Tyr Thr Pro Val Val Leu Pro Arg Ser Leu Asp
 225 230 235 240
 Phe Thr Glu Leu Asp Val Ala Ala Glu Lys Ile Asp Arg Phe Met Gln
 245 250 255
 Ala Val Thr Gly Trp Lys Thr Gly Cys Ser Leu Met Gly Ala Ser Val
 260 265 270
 Asp Ser Thr Leu Ala Phe Asn Thr Tyr Val His Phe Gln Gly Lys Met
 275 280 285
 Lys Gly Phe Ser Leu Leu Ala Glu Pro Gln Glu Phe Trp Val Asp Asn
 290 295 300
 Ser Thr Ser Val Ser Val Pro Met Leu Ser Gly Met Gly Thr Phe Gln
 305 310 315 320
 His Trp Ser Asp Ile Gln Asp Asn Phe Ser Val Thr Gln Val Pro Phe
 325 330 335
 Thr Glu Ser Ala Cys Leu Leu Leu Ile Gln Pro His Tyr Ala Ser Asp
 340 345 350
 Leu Asp Lys Val Glu Gly Leu Thr Phe Gln Gln Asn Ser Leu Asn Trp
 355 360 365
 Met Lys Lys Leu Ser Pro Arg Thr Ile His Leu Thr Met Pro Gln Leu
 370 375 380
 Val Leu Gln Gly Ser Tyr Asp Leu Gln Asp Leu Leu Ala Gln Ala Glu
 385 390 395 400
 Leu Pro Ala Ile Leu His Thr Glu Leu Asn Leu Gln Lys Leu Ser Asn
 405 410 415
 Asp Arg Ile Arg Val Gly Glu Val Leu Asn Ser Ile Phe Phe Glu Leu
 420 425 430
 Glu Ala Asp Glu Arg Glu Pro Thr Glu Ser Thr Gln Gln Leu Asn Lys

CEN5301WOPCTCORRECTEDSEQLIST.txt

435
 Pro Glu Val Leu Glu Val Thr Leu Asn Arg Pro Phe Leu Phe Ala Val
 450
 Tyr Asp Gln Ser Ala Thr Ala Leu His Phe Leu Gly Arg Val Ala Asn
 465
 Pro Leu Ser Thr Ala
 485

<210> 77
 <211> 133
 <212> PRT
 <213> Homo sapiens

<400> 77
 Met Val Asp Ala Phe Leu Gly Thr Trp Lys Leu Val Asp Ser Lys Asn
 1 5 10 15
 Phe Asp Asp Tyr Met Lys Ser Leu Gly Val Gly Phe Ala Thr Arg Gln
 20 25 30
 Val Ala Ser Met Thr Lys Pro Thr Ile Ile Glu Lys Asn Gly Asp
 35 40 45
 Ile Leu Thr Leu Lys Thr His Ser Thr Phe Lys Asn Thr Glu Ile Ser
 50 55 60
 Phe Lys Leu Gly Val Glu Phe Asp Glu Thr Thr Ala Asp Asp Arg Lys
 65 70 75 80
 Val Lys Ser Ile Val Thr Leu Asp Gly Gly Lys Leu Val His Leu Gln
 85 90 95
 Lys Trp Asp Gly Gln Glu Thr Thr Leu Val Arg Glu Leu Ile Asp Gly
 100 105 110
 Lys Leu Ile Leu Thr Leu Thr His Gly Thr Ala Val Cys Thr Arg Thr
 115 120 125
 Tyr Glu Lys Glu Ala
 130

<210> 78
 <211> 212
 <212> PRT
 <213> Homo sapiens

<400> 78
 Met Asn Ser Phe Ser Thr Ser Ala Phe Gly Pro Val Ala Phe Ser Leu
 1 5 10 15
 Gly Leu Leu Leu Val Leu Pro Ala Ala Phe Pro Ala Pro Val Pro Pro
 20 25 30
 Gly Glu Asp Ser Lys Asp Val Ala Ala Pro His Arg Gln Pro Leu Thr
 35 40 45
 Ser Ser Glu Arg Ile Asp Lys Gln Ile Arg Tyr Ile Leu Asp Gly Ile
 50 55 60
 Ser Ala Leu Arg Lys Glu Thr Cys Asn Lys Ser Asn Met Cys Glu Ser
 65 70 75 80
 Ser Lys Glu Ala Leu Ala Glu Asn Asn Leu Asn Leu Pro Lys Met Ala
 85 90 95
 Glu Lys Asp Gly Cys Phe Gln Ser Gly Phe Asn Glu Glu Thr Cys Leu
 100 105 110
 Val Lys Ile Ile Thr Gly Leu Leu Glu Phe Glu Val Tyr Leu Glu Tyr
 115 120 125
 Leu Gln Asn Arg Phe Glu Ser Ser Glu Glu Gln Ala Arg Ala Val Gln
 130 135 140
 Met Ser Thr Lys Val Leu Ile Gln Phe Leu Gln Lys Lys Ala Lys Asn
 145 150 155 160
 Leu Asp Ala Ile Thr Thr Pro Asp Pro Thr Thr Asn Ala Ser Leu Leu
 165 170 175
 Thr Lys Leu Gln Ala Gln Asn Gln Trp Leu Gln Asp Met Thr Thr His
 180 185 190
 Leu Ile Leu Arg Ser Phe Lys Glu Phe Leu Gln Ser Ser Leu Arg Ala
 195 200 205
 Leu Arg Gln Met

210

<210> 79
 <211> 154
 <212> PRT
 <213> Homo sapiens

<400> 79
 Met Gly Leu Ser Asp Gly Glu Trp Gln Leu Val Leu Asn Val Trp Gly
 1 5 10 15
 Lys Val Glu Ala Asp Ile Pro Gly His Gly Gln Glu Val Leu Ile Arg
 20 25 30
 Leu Phe Lys Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys
 35 40 45
 His Leu Lys Ser Glu Asp Glu Met Lys Ala Ser Glu Asp Leu Lys Lys
 50 55 60
 His Gly Ala Thr Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys
 65 70 75 80
 Gly His His Glu Ala Glu Ile Lys Pro Leu Ala Gln Ser His Ala Thr
 85 90 95
 Lys His Lys Ile Pro Val Lys Tyr Leu Glu Phe Ile Ser Glu Cys Ile
 100 105 110
 Ile Gln Val Leu Gln Ser Lys His Pro Gly Asp Phe Gly Ala Asp Ala
 115 120 125
 Gln Gly Ala Met Asn Lys Ala Leu Glu Leu Phe Arg Lys Asp Met Ala
 130 135 140
 Ser Asn Tyr Lys Glu Leu Gly Phe Gln Gly
 145 150

<210> 80
 <211> 496
 <212> PRT
 <213> Homo sapiens

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

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

      245      250      255
Asn Leu Leu Thr Met Met Ser Thr Ser Asn Ser Ala Lys Asp Pro Thr
      260      265      270
Val Ala Lys Glu Glu Gln Ile Ser Phe Arg Asp Cys Ala Glu Val Phe
      275      280      285
Lys Ser Gly His Thr Thr Asn Gly Ile Tyr Thr Leu Thr Phe Pro Asn
      290      295      300
Ser Thr Glu Glu Ile Lys Ala Tyr Cys Asp Met Glu Ala Gly Gly Gly
305      310      315      320
Gly Trp Thr Ile Ile Gln Arg Arg Glu Asp Gly Ser Val Asp Phe Gln
      325      330      335
Arg Thr Trp Lys Glu Tyr Lys Val Gly Phe Gly Asn Pro Ser Gly Glu
      340      345      350
Tyr Trp Leu Gly Asn Glu Phe Val Ser Gln Leu Thr Asn Gln Gln Arg
      355      360      365
Tyr Val Leu Lys Ile His Leu Lys Asp Trp Glu Gly Asn Glu Ala Tyr
      370      375      380
Ser Leu Tyr Glu His Phe Tyr Leu Ser Ser Glu Glu Leu Asn Tyr Arg
385      390      395      400
Ile His Leu Lys Gly Leu Thr Gly Thr Ala Gly Lys Ile Ser Ser Ile
      405      410      415
Ser Gln Pro Gly Asn Asp Phe Ser Thr Lys Asp Gly Asp Asn Asp Lys
      420      425      430
Cys Ile Cys Lys Cys Ser Gln Met Leu Thr Gly Gly Trp Trp Phe Asp
      435      440      445
Ala Cys Gly Pro Ser Asn Leu Asn Gly Met Tyr Tyr Pro Gln Arg Gln
450      455      460
Asn Thr Asn Lys Phe Asn Gly Ile Lys Trp Tyr Tyr Trp Lys Gly Ser
465      470      475      480
Gly Tyr Ser Leu Lys Ala Thr Thr Met Met Ile Arg Pro Ala Asp Phe
      485      490      495

```

<210> 81

<211> 477

<212> PRT

<213> Homo sapiens

<400> 81

```

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

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

225      230      235      240
His Ser Leu Thr Asp Leu Thr Arg Phe Arg Leu Ser Gln Asp Asp Ile
      245      250      255
Asn Gly Ile Gln Ser Leu Tyr Gly Pro Pro Asp Ser Pro Glu Thr
      260      265      270
Pro Leu Val Pro Thr Glu Pro Val Pro Pro Glu Pro Gly Thr Pro Ala
      275      280      285
Asn Cys Asp Pro Ala Leu Ser Phe Asp Ala Val Ser Thr Leu Arg Gly
      290      295      300
Glu Ile Leu Ile Phe Lys Asp Arg His Phe Trp Arg Lys Ser Leu Arg
305      310      315      320
Lys Leu Glu Pro Glu Leu His Leu Ile Ser Ser Phe Trp Pro Ser Leu
      325      330      335
Pro Ser Gly Val Asp Ala Ala Tyr Glu Val Thr Ser Lys Asp Leu Val
      340      345      350
Phe Ile Phe Lys Gly Asn Gln Phe Trp Ala Ile Arg Gly Asn Glu Val
      355      360      365
Arg Ala Gly Tyr Pro Arg Gly Ile His Thr Leu Gly Phe Pro Pro Thr
      370      375      380
Val Arg Lys Ile Asp Ala Ile Ser Asp Lys Glu Lys Asn Lys Thr
385      390      395      400
Tyr Phe Phe Val Glu Asp Lys Tyr Trp Arg Phe Asp Glu Lys Arg Asn
      405      410      415
Ser Met Glu Pro Gly Phe Pro Lys Gln Ile Ala Glu Asp Phe Pro Gly
      420      425      430
Ile Asp Ser Lys Ile Asp Ala Val Phe Glu Glu Phe Gly Phe Phe Tyr
      435      440      445
Phe Phe Thr Gly Ser Ser Gln Leu Glu Phe Asp Pro Asn Ala Lys Lys
      450      455      460
Val Thr His Thr Leu Lys Ser Asn Ser Trp Leu Asn Cys
465      470      475

```

<210> 82
 <211> 120
 <212> PRT
 <213> Homo sapiens

```

<400> 82
Met Ala Ser His Arg Leu Leu Leu Leu Cys Leu Ala Gly Leu Val Phe
 1      5      10      15
Val Ser Glu Ala Gly Pro Thr Gly Thr Gly Glu Ser Lys Cys Pro Leu
      20      25      30
Met Val Lys Val Leu Asp Ala Val Arg Gly Ser Pro Ala Ile Asn Val
      35      40      45
Ala Val His Val Phe Arg Lys Ala Ala Asp Asp Thr Trp Glu Pro Phe
      50      55      60
Ala Ser Gly Lys Thr Ser Glu Ser Gly Glu Leu His Gly Leu Thr Thr
65      70      75      80
Glu Glu Glu Phe Val Glu Gly Ile Tyr Lys Val Glu Ile Asp Thr Lys
      85      90      95
Ser Tyr Trp Lys Ala Leu Gly Ile Ser Pro Phe His Glu His Ala Glu
      100      105      110
Val Val Phe Thr Ala Asn Asp Ser
      115      120

```

<210> 83
 <211> 222
 <212> PRT
 <213> Homo sapiens

```

<400> 83
Met Ala Glu Lys Pro Lys Leu His Tyr Phe Asn Ala Arg Gly Arg Met
 1      5      10      15
Glu Ser Thr Arg Trp Leu Leu Ala Ala Ala Gly Val Glu Phe Glu Glu
      20      25      30
Lys Phe Ile Lys Ser Ala Glu Asp Leu Asp Lys Leu Arg Asn Asp Gly

```

CEN5301WOPCTCORRECTEDSEQLIST.txt

```

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

```