



(22) Date de dépôt/Filing Date: 2004/07/27
(41) Mise à la disp. pub./Open to Public Insp.: 2005/02/10
(45) Date de délivrance/Issue Date: 2015/04/28
(62) Demande originale/Original Application: 2 533 938
(30) Priorité/Priority: 2003/07/30 (US60/491,151)

(51) Cl.Int./Int.Cl. *C12N 15/53* (2006.01),
C12N 1/21 (2006.01), *C12N 15/63* (2006.01),
C12N 9/02 (2006.01), *C12P 7/40* (2006.01),
C12P 7/58 (2006.01)

(72) Inventeurs/Inventors:
CALDWELL, ROBERT M., US;
RASHID, HARUNUR M., US;
VALLE, FERNANDO, US

(73) Propriétaire/Owner:
GENENCOR INTERNATIONAL, INC., US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : OXYDOREDUCTASES MULTIMERES
(54) Title: MULTIMERIC OXYDOREDUCTASES

(57) **Abrégé/Abstract:**

The present invention concerns multimeric oxidoreductase complexes which function in the enzymatic conversion of a carbon substrate, said complexes having a dehydrogenase subunit and a cytochrome C subunit. The invention further relates to polynucleotides coding for the multimeric complexes and methods of use thereof.

ABSTRACT

The present invention concerns multimeric oxidoreductase complexes which function in the enzymatic conversion of a carbon substrate, said complexes having a dehydrogenase subunit and a cytochrome C subunit. The invention further relates to polynucleotides coding for the multimeric complexes and methods of use thereof.

MULTIMERIC OXIDOREDUCTASES

1. FIELD OF THE INVENTION

The present invention relates to isolated naturally occurring and recombinant multimeric oxidoreductase complexes. More specifically the invention relates to the identification and use of multimeric oxidoreductase complexes isolated from *Pantoea citrea* wherein the complexes include a cytochrome C subunit and a dehydrogenase subunit.

2. BACKGROUND OF THE INVENTION

The synthesis and manufacture of L-ascorbic acid (AsA, vitamin C) has received considerable attention due to its relatively large market volume and high value as a vitamin and antioxidant. A chemical route, the Reichstein-Grussner method, from glucose to AsA, was first disclosed in 1934 (*Helv. Chim. Acta* 17:311-328). More recently bioconversion methods for the production of AsA intermediates have been disclosed and reference is made to Lazarus et al. (1989), "Vitamin C: Bioconversion via a Recombinant DNA Approach", *GENETICS AND MOLECULAR BIOLOGY OF INDUSTRIAL MICROORGANISMS*, American Society for Microbiology, Washington D.C. edited by C.L. Hershberger; Crawford et al., (1980) *Advances in Carbohydrate Chemistry and Biochemistry* 37: 79 - 155 (1980); Anderson et al., (1985) *Sci.* 230: 144 - 149; and Sugisawa et al. (1990) *Agric. Biol. Chem.* 54:1201 - 1209.

A typical process for the manufacture of AsA is shown in Figure 15. In general, the process begins with a metabolite used by a microorganism, e.g. D-glucose. Through enzymatic conversions, which may include D-glucose dehydrogenase, D-gluconate dehydrogenase, and 2-keto-D-gluconate dehydrogenase, the D-glucose undergoes a series of oxidative steps to yield 2,5-diketo-D-gluconate (2,5-DKG). Further the 2,5-DKG

- 2 -

may be reduced to 2- keto-L-gulonic acid (2-KLG). This process may be carried out in microorganisms such as *Gluconobacter*, *Acetobacter*, *Erwinia* or *Pantoea*. Reference is made to various U.S. Patents disclosing parts of this overall conversion such as U.S. Patent Nos. 3,790,444; 3,922,194; 3,959,076; 3,998,697; 4,245,049 and 5,008,193.

5 Because of the commercial market for AsA, AsA intermediates independent of AsA, have become a material of economic and industrial importance and for that reason it would be desirable to increase microbial efficiency for enzymatic conversion of carbon substrates into AsA intermediates.

Compared to other bacterial organisms, the gram-negative Enterobacteria,
10 *Pantoea citrea* has the ability to efficiently convert glucose and other sugars into different aldo and keto-sugar derivatives and particularly into the AsA intermediates 2,5-DKG and 2-KGL. In this invention, the genome of *Pantoea citrea* was analyzed to determine if there were unique properties of the microorganism that contributed to efficient sugar conversion. Analysis of the genome revealed that while the *Pantoea* genome is similar to
15 other Enterobacteria, such as *Salmonella*, *Klebsiella* and *E. coli*, in certain respects, it contains a number of different genes, which provide additional sugar metabolism capabilities.

This invention is directed to the discovery that the *Pantoea citrea* genome includes a family of genes that code for membrane bound three-component
20 oxidoreductase complexes. Specifically, it was discovered that the *P. citrea* genome includes 19 operons that code for membrane bound three-component oxidoreductase complexes and each of these complexes include a cytochrome C homologue subunit and a subunit having dehydrogenase activity. This is in contrast to other known microbial genomes, such as *Bacillus subtilis*, *E. coli* and *Pseudomonas aeruginosa*. *B. subtilis* and
25 *E. coli* are not known to include multimeric oxidoreductase enzyme complexes containing cytochrome C homologs and *P. aeruginosa* is known to comprise one three-component multimeric complex containing a cytochrome C homologue.

30

3. SUMMARY OF THE INVENTION

In a first aspect, the invention is directed to isolated polynucleotides encoding cytochrome C (CytC) proteins and the CytC proteins encoded by said polynucleotides. Preferred CytC proteins include the amino acid sequences illustrated in Figures 4A - 4B
35 and 5A - 5B.

- 3 -

In a second aspect, the invention is directed to isolated polynucleotides encoding dehydrogenases and the dehydrogenase proteins encoded by said polynucleotides. Preferred dehydrogenases include the amino acid sequences illustrated in Figures 2A - 2C and 3A - 3B.

5 In a third aspect, the invention is related to polynucleotides encoding multimeric oxidoreductase complexes, wherein the complex comprises a CytC subunit and a dehydrogenase subunit, and the multimeric complexes encoded by said polynucleotides.

In a fourth aspect, the invention is related to polynucleotides encoding multimeric oxidoreductase complexes comprising an operon including three subunits said subunits
10 including a CytC subunit, a dehydrogenase subunit and a third subunit designated the gamma subunit.

In a fifth aspect, the invention is directed to an operon encoding an oxidoreductase complex, said operon comprising three coding regions, wherein the coding regions are the regions illustrated by the 19 operons in Figures 1A and 1B. In one
15 embodiment of this aspect, the operon comprises the open reading frames (orfs) of orf 3653, orf 3652 and orf 3651. In a second embodiment of this aspect, the operon comprises the orfs of orf 2418, orf 2419 and orf 2420. In a third embodiment of this aspect, the operon comprises the orfs of orf 1840, orf 1841 and orf 1842. In a fourth embodiment of this aspect, the operon comprises the orfs of orf 2035, orf 2036 and orf
20 2037. In a fifth embodiment of this aspect, the operon comprises the orfs of orf 3687, orf 3688 and orf 3689.

In a sixth aspect, the invention concerns a vector comprising a polynucleotide encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex as defined herein. Further the invention is directed to host cells transformed with
25 a vector of the invention. Preferred host cells include *Pantoea*, *Klebsiella* and *E. coli* cells.

In a seventh aspect, the invention is directed to a method of enhancing enzymatic oxidation of a carbon substrate in a host cell comprising transforming the host cell with a polynucleotide encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex as defined herein to obtain an altered host cell, wherein the
30 polynucleotide is expressed; culturing the altered host cell under suitable culture conditions and allowing the oxidation of the substrate wherein said oxidation is enhanced compared to a corresponding unaltered host cell. In a preferred embodiment of this aspect the substrate is a polyol.

In an eighth aspect, the invention is directed to a method of enhancing enzymatic
35 reduction of a carbon substrate in a host cell comprising transforming the host cell with a

- 4 -

polynucleotide encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex as defined herein to obtain an altered host cell, wherein the polynucleotide is expressed; culturing the altered host cell under suitable culture conditions and allowing the reduction of the substrate wherein said reduction is enhanced compared to a corresponding unaltered host cell. In a preferred embodiment of this aspect the substrate is a polyol.

In a ninth aspect, the invention is directed to a method for enhancing the production of an AsA intermediate in a host cell. One embodiment of this aspect comprises introducing into a host cell a polynucleotide encoding a multimeric oxidoreductase complex including at least two subunits as defined herein to obtain an altered host cell, wherein the polynucleotide is expressed; culturing the altered host cell under suitable culture conditions and allowing the production of an AsA intermediate. In a second embodiment of this aspect, the AsA intermediate is preferably 2-KDG or 2,5-DKG. In another embodiment of this aspect, the polynucleotide encodes a multimeric oxidoreductase complex selected from the group consisting of

- a) SEQ ID NOs. 41, 5, and 23
- b) SEQ ID NOs. 42, 4 and 22;
- c) SEQ ID NOs. 39, 6 and 24;
- d) SEQ ID NOs. 40, 3 and 29;
- e) SEQ ID NOs. 57, 19 and 38;
- f) sequences having at least 95%, 96%, 97%, 98% and 99% sequence identity to each of the sequences of a) - e) above; and
- g) combinations of each of the sequences of a) - f) above.

25

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show a schematic representation of the 19 operons identified and sequenced according to the invention. Each operon is characterized by at least three open reading frames (Orfs) generally designated as the gamma subunit (shaded symbol), the alpha subunit (blank symbol) and the cytochrome C (CytC) subunit (stippled symbol). The alpha subunit has dehydrogenase activity. In figure 1A the order of the gamma, alpha and CytC subunits is preserved in the operon and this group of 10 operons is referred to herein as family 1. In figure 1B the order of the gamma, alpha and CytC subunits is not always preserved and this group of operons is referred to herein as

- 5 -

family 2. The bold arrows indicate flanking genes in the same region of the respective open reading frame (orf).

- Orf 764 corresponds to the amino acid sequence of SEQ ID NO: 46;
Orf 765 corresponds to the amino acid sequence of SEQ ID NO: 1;
5 Orf 766 corresponds to the amino acid sequence of SEQ ID NO: 20;
Orf 1748 corresponds to the amino acid sequence of SEQ ID NO: 48;
Orf 1749 corresponds to the amino acid sequence of SEQ ID NO: 8;
Orf 1750 corresponds to the amino acid sequence of SEQ ID NO: 27;
Orf 1842 corresponds to the amino acid sequence of SEQ ID NO: 41;
10 Orf 1841 corresponds to the amino acid sequence of SEQ ID NO: 5;
Orf 1840 corresponds to the amino acid sequence of SEQ ID NO: 23;
Orf 1951 corresponds to the amino acid sequence of SEQ ID NO: 44;
Orf 1952 corresponds to the amino acid sequence of SEQ ID NO: 7;
Orf 1953 corresponds to the amino acid sequence of SEQ ID NO: 28;
15 Orf 2037 corresponds to the amino acid sequence of SEQ ID NO: 42;
Orf 2036 corresponds to the amino acid sequence of SEQ ID NO: 4;
Orf 2035 corresponds to the amino acid sequence of SEQ ID NO: 22;
Orf 2634 corresponds to the amino acid sequence of SEQ ID NO: 45;
Orf 2633 corresponds to the amino acid sequence of SEQ ID NO: 9;
20 Orf 2632 corresponds to the amino acid sequence of SEQ ID NO: 26;
Orf 2910 corresponds to the amino acid sequence of SEQ ID NO: 43;
Orf 2909 corresponds to the amino acid sequence of SEQ ID NO: 2;
Orf 2908 corresponds to the amino acid sequence of SEQ ID NO: 21;
Orf 3051 corresponds to the amino acid sequence of SEQ ID NO: 47;
25 Orf 3052 corresponds to the amino acid sequence of SEQ ID NO: 10;
Orf 3053 corresponds to the amino acid sequence of SEQ ID NO: 25;
Orf 3653 corresponds to the amino acid sequence of SEQ ID NO: 39;
Orf 3652 corresponds to the amino acid sequence of SEQ ID NO: 6;
Orf 3651 corresponds to the amino acid sequence of SEQ ID NO: 24;
30 Orf 3689 corresponds to the amino acid sequence of SEQ ID NO: 40;
Orf 3688 corresponds to the amino acid sequence of SEQ ID NO: 3;
Orf 3687 corresponds to the amino acid sequence of SEQ ID NO: 29;
Orf 1219 corresponds to the amino acid sequence of SEQ ID NO: 52;
Orf 1220 corresponds to the amino acid sequence of SEQ ID NO: 18;
35 Orf 1221 corresponds to the amino acid sequence of SEQ ID NO: 34;

- 6 -

Orf 1982 corresponds to the amino acid sequence of SEQ ID NO: 55;
 Orf 1983 corresponds to the amino acid sequence of SEQ ID NO: 14;
 Orf 1984 corresponds to the amino acid sequence of SEQ ID NO: 35;
 Orf 1957 corresponds to the amino acid sequence of SEQ ID NO: 53;
 5 Orf 1956 corresponds to the amino acid sequence of SEQ ID NO: 31;
 Orf 1955 corresponds to the amino acid sequence of SEQ ID NO: 16;
 Orf 2418 corresponds to the amino acid sequence of SEQ ID NO: 57;
 Orf 2419 corresponds to the amino acid sequence of SEQ ID NO: 19;
 Orf 2420 corresponds to the amino acid sequence of SEQ ID NO: 38;
 10 Orf 2448 corresponds to the amino acid sequence of SEQ ID NO: 12;
 Orf 2447 corresponds to the amino acid sequence of SEQ ID NO: 50;
 Orf 2446 corresponds to the amino acid sequence of SEQ ID NO: 32;
 Orf 3043 corresponds to the amino acid sequence of SEQ ID NO: 54;
 Orf 3042 corresponds to the amino acid sequence of SEQ ID NO: 13;
 15 Orf 3041 corresponds to the amino acid sequence of SEQ ID NO: 37;
 Orf 3397 corresponds to the amino acid sequence of SEQ ID NO: 58;
 Orf 3398 corresponds to the amino acid sequence of SEQ ID NO: 56;
 Orf 3399 corresponds to the amino acid sequence of SEQ ID NO: 15;
 Orf 3400 corresponds to the amino acid sequence of SEQ ID NO: 36;
 20 Orf 3676 corresponds to the amino acid sequence of SEQ ID NO: 51;
 Orf 3675 corresponds to the amino acid sequence of SEQ ID NO: 17;
 Orf 3674 corresponds to the amino acid sequence of SEQ ID NO: 33;
 Orf 3820 corresponds to the amino acid sequence of SEQ ID NO: 30;
 Orf 3819 corresponds to the amino acid sequence of SEQ ID NO: 49; and
 25 Orf 3818 corresponds to the amino acid sequence of SEQ ID NO: 11.

Figures 2A - 2C illustrate the amino acid sequences of the alpha subunits corresponding to the family 1 operons including SEQ ID NOs:1 - 10.

Figures 3A - 3B illustrate the amino acid sequences of the alpha subunits
 30 corresponding to the family 2 operons including SEQ ID NOs: 11 - 19.

Figures 4A - 4B illustrate the amino acid sequences of the CytC subunits corresponding to family 1 operons, including SEQ ID NOs: 20 - 29.

Figures 5A - 5B illustrate the amino acid sequences of the CytC subunits corresponding to family 2 operons, including SEQ ID NOs: 30 - 38.

- 7 -

Figures 6A - 6B illustrate the amino acid sequences of the gamma subunit corresponding to family 1 operons, including SEQ ID NOs: 39 - 48.

Figure 7 illustrates the amino acid sequences of the gamma subunit corresponding to family 2 operons, including SEQ ID NOs: 49 - 57.

5 Figure 8 illustrates the amino acid sequence of Orf 3397 corresponding to SEQ ID NO: 58.

Figures 9A - 9J illustrate the nucleic acid sequences (SEQ ID NOs. 59 - 68) encoding the amino acid sequences of the family 1 alpha subunits illustrated in Figures 2A - 2C.

10 Figures 10A - 10H illustrate the nucleic acid sequences (SEQ ID NOs. 69 - 77) encoding the amino acid sequences of the family 2 alpha subunits illustrated in Figures 3A - 3B.

Figures 11A - 11H illustrate the nucleic acid sequences (SEQ ID NOs 78 - 87) encoding the amino acid sequences of the family 1 CytC subunits illustrated in Figures 15 4A - 4B.

Figures 12A - 12H illustrate the nucleic acid sequences (SEQ ID NOs. 88 - 96) encoding the amino acid sequences of the family 2 CytC subunits illustrated in Figures 5A - 5B.

Figures 13A - 13E illustrate the nucleic acid sequences (SEQ ID NOs. 97 - 106) encoding the amino acid sequences of the family 1 gamma subunits illustrated in Figures 20 6A - 6B.

Figures 14A - 14C illustrate the nucleic acid sequences (SEQ ID NOs. 107 - 115) encoding the amino acid sequences of the family 1 alpha subunits illustrated in Figure 7.

Figure 15 is a schematic representation of a pathway to AsA intermediates. In 25 this representation, the enzyme that converts glucose (GL) to gluconate (GA) is glucose dehydrogenase (GDH); the oxidative enzyme that converts GA to 2-KDG is gluconate dehydrogenase (GADH); the oxidative enzyme that converts 2KDG to 2,5-DKG is 2-keto-D-gluconate dehydrogenase (KDGDH) and the reducing enzyme that converts 2, 5-DKG to 2-KLG is 2,5-diketo-D-gluconic acid reductase (DKGR). The 2-KLG may then be 30 recovered and chemically converted to L-Ascorbic Acid (AsA).

Figures 16A and 16B depict the three heme binding sites found in the CytC proteins illustrated in figures 4 and 5 and corresponding to Orfs 766, 1840, 1953, 2035, 2632, 2908, 3053, 3651, 3687, 1750, 3820, 1956, 1984, 2420, 2446, 3041, 3400, 3674 and 1221.

- 8 -

5. DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology and biochemistry, which are within the skill of one in the art. Such techniques are explained fully in the literature, such as MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. (Sambrook, J. et al., 1989) Cold Spring Harbor Laboratory Press; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel et al., eds., 1987 and annual updates); OLIGONUCLEOTIDE SYNTHESIS (M. J. Gait ed., 1984); PCR: THE POLYMERASE CHAIN REACTION (Mullis et al., eds., 1994); MANUAL OF INDUSTRIAL MICROBIOLOGY AND BIOTECHNOLOGY, 2nd Ed. (A.L. Demain et al., eds 1999); MANUAL OF METHODS FOR GENERAL BACTERIOLOGY (P. Gerhardt et al., eds) pp. 210 – 213, American Society for Microbiology, Washington D.C., and BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY, 2nd Ed. (T. D. Brock, 1989) Sinauer Associates, Inc. Sunderland, MA.

It is to be understood that this invention is not limited to the particular methodology, protocols and reagents described, as these may vary.

All patent and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

The headings provided herein are not limitations of the various aspects or embodiments of the invention, which can be had by reference to the specification as a whole.

A. Definitions.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994) and Hale & Marham, THE HARPER DICTIONARY OF BIOLOGY, Harper Perennial, New York (1991) provide one of skill with general dictionaries of many of the terms used in this invention.

The following abbreviations apply as used herein: glucose (GL); D-gluconate (GA); 2-keto-D-gluconate (2-KDG); 2,5-diketo-D-gluconate (2,5-DKG), 2-keto-L-gulonic acid (2-KLG), L-idonic acid (IA), 5-keto-D-gluconate (5-KDG), ascorbic acid (AsA), glucose dehydrogenase (GDH), gluconate dehydrogenase (GADH), 2-keto-D-gluconate dehydrogenase (2-KDGDH), and 2,5-diketo-D-gluconate reductase (DKGR).

- 9 -

The term "AsA intermediate" encompasses any product in the pathway to AsA including but not limited to GA, 2-KDG, 2,5-DKG, 5-KDG, IA and 2-KLG.

Four stereoisomers of AsA are possible: L-ascorbic acid (AsA), D-araboascorbic acid (erythorbic acid), which shows vitamin C activity, L-araboascorbic acid, and D-xyloascorbic acid.

As used herein "GDH" or "a protein having GDH activity" refers to a protein which is capable of catalyzing the conversion of glucose (GL) stereoselectively to (GA).

The term "GADH" or "a protein having GADH activity" refers to a protein which is capable of catalyzing the conversion of GA stereoselectively to 2-KDG. For example, the multimeric oxidoreductase complexes comprising the operons selected from the group of

- a) orfs 1842, 1841 and 1840;
- b) orfs 2037, 2036 and 2035;
- c) orfs 3653, 3652 and 3651; and
- d) orfs 3689, 3688 and 3687

have GADH activity.

The term "2-KDGDH" or "a protein having 2-KDGDH activity" refers to a protein which is capable of catalyzing the conversion of 2-KDG stereoselectively to 2,5-DKG, and 2,5-diketo-D-gluconate reductase (DKGR) refers to a protein capable of catalyzing the conversion of 2,5-DKG stereoselectively to 2-KLG.

An "oxidoreductase protein" as used herein broadly refers to an oxidation-reduction (redox) enzyme or electron-transferring protein that is involved in the transfer of electrons from an electron donor (the reducing agent or reductant) to an electron acceptor (the oxidizing agent or oxidant) such as molecular oxygen.

There are generally five different groups of oxidoreductase proteins:

a) Pyridine-linked dehydrogenases, which require either nicotinamide adeninedinucleotide (NAD) or nicotine adeninedinucleotide phosphate (NADP) as a coenzyme and catalyze the reversible transfer of electrons from a donor to the coenzyme to form either NADH or NADPH;

b) flavin-linked dehydrogenases, which contain flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as a prosthetic group;

c) iron-sulfur proteins, which contain two to eight atoms of iron and an equal number of acid-labile sulfur atoms wherein the iron atoms undergo Fe(II) -Fe(III) transitions;

d) cytochromes which contain iron-porphyrin prosthetic groups; and

e) ubiquinone also known as coenzyme Q, a lipid soluble coenzyme.

- 10 -

Examples of pyridine-linked dehydrogenase systems include NAD linked glyceraldehyde 3-phosphate dehydrogenases; lactate dehydrogenases; glycerol 3-phosphate dehydrogenases; glutamate dehydrogenases and ethanol dehydrogenases, and NADP linked glucose 6-phosphate dehydrogenases and isocitrate dehydrogenases.

5 Examples of flavin-linked dehydrogenase systems include NADH dehydrogenases, which contain FMN and catalyze the transfer of electrons from NADH; succinate dehydrogenases; dihydrolipoyl dehydrogenases; α -ketoglutarate dehydrogenases; acyl-CoA dehydrogenases; glucose oxidases and aldehyde oxidases. Examples of iron-sulfur proteins include NADH dehydrogenases; succinate dehydrogenases and ferredoxins. At

10 least five different types of cytochromes have been identified based upon absorption spectral data and these include Cytochromes a, a_3 , b, c, and c_1 .

A "multimeric oxidoreductase complex" as used herein means at least two polypeptide subunits operably linked to one another having oxidoreductase activity. In one preferred embodiment, the multimeric oxidoreductase complex comprises three

15 subunits organized in a single transcriptional unit known as an operon. A multimeric oxidoreductase complex according to the invention encompasses an oxidoreductase enzyme or protein within any one of the five oxidoreductase groups referred to above.

The term "subunit" as used herein generally refers to a discrete continuous part of an amino acid sequence that is equated with a particular function and is used

20 interchangeably with the term domain.

The term "cytochrome C (CytC)" or "CytC protein" as used herein refers to an oxidoreductase having one or several heme c groups bound to the protein by one, or more commonly two, thioether bonds involving sulphydryl groups of cysteine residues. The fifth heme iron ligand is always provided by a histidine residue (Pettigrew et al.

25 (1987) CYTOCHROMES C. BIOLOGICAL ASPECTS, Springer Verlag, Berlin; Moore et al. (1990) CYTOCHROMES C: EVOLUTIONARY, STRUCTURAL AND PHYSIOCHEMICAL ASPECTS, Springer Verlag, Berlin; and Ambler (1991), *Biochim. Biophys. Acta* 1058:42 – 47).

The term "CytC subunit" refers to a component of a multimeric oxidoreductase as referred to above, which has the structure and function of a CytC as defined above.

30 The term "dehydrogenase" as used herein refers to an enzyme that catalyzes an oxidoreductase reaction involving removal of hydrogen from one substrate and its transfer to another molecule, usually to a coenzyme or cofactor, such as NAD, NADP and FAD. Preferred dehydrogenases according to the meaning herein fall under the group designated pyridine-linked dehydrogenases which require either NAD or NADP as

35 a coenzyme and catalyze the reversible transfer of electrons from a donor to a coenzyme

- 11 -

to form either NADH or NADPH. and flavin-linked dehydrogenases, which contain flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as a cofactor.

The term "dehydrogenase subunit" used interchangeably with the term "alpha subunit" refers to a component of a multimeric oxidoreductase as referred to above, which has the structure and function of a dehydrogenase as defined above.

As used herein, "nucleic acid" refers to a nucleotide or polynucleotide sequence, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be double-stranded or single-stranded, whether representing the sense or antisense strand.

As used herein the term "polypeptide" refers to a compound made up of amino acid residues linked by peptide bonds. "Amino acid" refers to all naturally occurring L- α -amino acids and includes norleucine, ornithine and homocysteine. The amino acids are identified by either the single-letter or three-letter designations. The term "protein" as used herein is synonymous with the term "polypeptide" and may also refer to two or more polypeptides. Thus the terms "protein", "peptide" and "polypeptide" are used interchangeably. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein may be produced.

A polynucleotide or polypeptide having a certain percentage (for example 80%, 85%, 90% or 95%) of sequence identity to another sequence means that, when aligned, that percentage of nucleotide bases or amino acid residues in a candidate sequence are identical with the nucleotide bases or amino acid residues of the reference sequence.

The term "operon" as used herein means a cluster of two or more structural genes which are transcribed as a single transcriptional unit from a common promoter. The genes of the operon code for proteins with related metabolic function. The genes comprising an operon may be adjacent genes.

The term "gene" means the segment of DNA involved in producing a polypeptide chain, that may or may not include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or leader sequences and 3' UTR or "trailer" sequences as well as intervening sequences (introns) between individual coding segments (exons).

The term "open reading frame (orf)" is defined herein as a region of a gene that contains the coding sequence for a protein.

A "vector" refers to a nucleic acid construct designed for transfer of nucleic acid sequences into host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commonly available. Selection of appropriate

- 12 -

expression vectors is within the knowledge of those having skill in the art. An expression vector is generally generated recombinantly or synthetically with a series of specific nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector may be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus or nucleic acid fragment. In the present specification "plasmid" and "vector" are often used interchangeably. Typically a vector includes among other sequences, a nucleic sequence to be transcribed and a promoter.

A "promoter" as used herein refers to a nucleic acid sequence that functions to direct transcription of a downstream gene or genes. The promoter will generally be appropriate to the host cell in which the desired gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences are necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include but are not limited to promoter sequences, ribosomal binding sites, transcriptional start and stop sequences translational start and stop sequence and enhancer or activator sequences.

The term "operably linked" means the nucleic acid is placed in a functional relationship with another nucleic acid sequence. For example, a promoter is operably linked to a coding sequence if it affects the transcription of the sequence. Generally, "operably linked" means the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phrase. However, enhancers do not have to be contiguous. Linking may be accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The terms "isolated" and "purified" as used herein refer to any enzyme, nucleic acid, protein, peptide or cofactor that is removed from at least one component which it is naturally associated.

"Chromosomal integration" is a process whereby an introduced nucleic acid is incorporated into a host chromosome. The process preferably takes place by homologous recombination.

"Recombinant host cells", "host cells", "cells" and "cell culture" are used interchangeably to designate individual cells, cell lines, cell cultures and harvested cells which have been or are intended to be transformed with a vector or construct of the invention. The term also includes the progeny of the cells originally receiving the vector or construct.

- 13 -

As used herein the term "recombinant" refers to a host cell that has a modification of its genome, e.g., as by the addition of nucleic acids not naturally occurring in the host organism or by a modification of nucleic acid naturally occurring in the host cell.

An "altered host cell" refers to a host cell comprising one or more introduced polynucleotides encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex according to the invention.

A "corresponding unaltered host cell" is a host cell which does not include the introduced polynucleotide encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex and is generally compared to an altered bacteria host when both are grown under essentially the same culture conditions.

The term "enhanced" refers to a level of activity, production or accumulation of an enzyme, substrate or product in an altered host cell compared to a corresponding unaltered host cell.

The term "polyol" means an alcohol molecule with numerous hydroxyl radicals. A keto-polyol derivative is a polyol that includes one or more keto groups in the polyol molecule. Non-limiting examples of polyols include glucose, gluconate, arabinose, arabitol, xylose, xylulose, xylitol, sucrose, sorbose, fructose, mannose, mannitol, meso-erythritol, L-erythrulose, idonate, cellobiose, lactose, idose, galactose, ribose, altose, maltose, erythrose, sorbitol, glycerol, 2-KDG, 5-keto-D-gluconic acid, 2-KLG, methanol, and erythorbic acid.

A "polyol dehydrogenase" is an enzyme which catalyzes an oxidoreductase reaction and uses a polyol substrate.

The term "carbon substrate" encompasses carbohydrates, including but not limited to glucose, gulose, lactose, sorbose, fructose, idose, galactose, mannose, ribose, xylose, arabinose, glycerol and dihydroxyacetone in either D or L form, or a combination of carbohydrates, such as glucose and fructose, and 6-carbon sugar acids, such as but not limited to 2-KLG, GA, idonic acid, 6-phosphogluconate, 2-KDG, 5-keto-D-gluconic acid, 2, 3-L-diketogulonic acid, dehydroascorbic acid, erythorbic acid and mannonic acid. In some preferred embodiments the carbon substrate is a hexose or a pentose. The terms polyol and carbon substrate may be used interchangeably herein.

A "desired product" herein refers to a compound which is enzymatically derived from the carbon substrate. In particularly preferred embodiments, the desired product is an organic acid (i.e. gluconic acid and other AsA intermediates, succinic acid, citric acid, lactic acid).

- 14 -

As used herein, the family "*Enterobacteriaceae*" refers to bacterial strains having the general characteristics of being gram negative and being facultatively anaerobic. One embodiment of preferred *Enterobacteriaceae* strains are those that are able to produce 2,5-diketo-D-gluconic acid from D-glucose solutions. Included in the family of *Enterobacteriaceae* are the genus *Erwinia*, *Enterobacter*, *Gluconobacter*, *Klebsiella*, *Escherichia* and *Pantoea*.

In the present invention, a preferred *Enterobacteriaceae* strain is a *Pantoea* species. *Pantoea* include *P. agglomerans*, *P. dispersa*, *P. punctata*, *P. citrea*, *P. terrea*, *P. ananas* and *P. stewartii*. Particularly preferred are strains of *Pantoea citrea*. *Pantoea citrea* can be obtained from ATCC (Manassas, VA) having ATCC accession number 39140, for example. *Pantoea citrea* has sometimes been referred to as *Erwinia citreus* or *Acetobacter cerinus*. Thus, it is intended that the genus *Pantoea* include species that have been reclassified, including but not limited to *Erwinia citreus* or *Acetobacter cerinus*.

In another embodiment, a preferred bacterial host strain is a strain from the *Pseudomonadaceae* family. These bacterial strains are gram-negative and generally non-sporing. Included in this family are *P. aeruginosa*, *P. alcaligenes*, *P. fluorescens*, *P. denitrificans*, *P. putida*, *P. species* (sp), *P. syringae*, *P. oleovorans*, *P. mendocina*, *P. pseudoalcaligenes*.

Other preferred bacterial host cells are *Bacillus* cells. *Bacillus* includes all members known to those of skill in the art, including but not limited to *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alcalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurnas*, *B. megaterum*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*.

The term "heterologous" as used herein refers to nucleic acid or amino acid sequences not naturally occurring in the host cell. As used herein, the term "endogenous" refers to a nucleic acid naturally occurring in the host.

The term "expression" as used herein refers to the process by which a polypeptide or protein is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation. The term "overexpression" means an increased number of copies of the same gene product in a host cell.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transformation", "transduction" or "transfection" and includes reference to the incorporation of a nucleic acid sequence into a prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (for example, chromosome, plasmid or

- 15 -

mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA or DNA).

The term "culturing" refers to fermentative bioconversion of a carbon substrate to a desired product within a culture vessel. In particularly preferred embodiments, culturing involves the growth of a bacterial organism under suitable conditions for the production of the desired products.

The term "enzymatic conversion" refers to the modification of a carbon substrate to a desired product by contacting the carbon substrate with an oxidoreductase enzyme complex or dehydrogenase enzyme encompassed by the invention. In some embodiments, contacting comprises exposing the carbon substrate to an altered bacterial host cell that expresses the multimeric oxidoreductase complex. In some embodiments the enzymatic conversion of a carbon substrate is an oxidation and may be part of an oxidative pathway yielding a desired product. In a nonlimiting example, gluconate is an oxidative product of the enzymatic conversion of glucose. In a further nonlimiting example the enzymatic conversion of glucose to 2-KDG and 2,5-DKG includes two or three enzymes in an oxidative pathway. In some embodiments the enzymatic conversion of a carbon substrate is a reduction.

It is well known in the art that many compounds mentioned in the present specification, such as gluconate may exist in a variety of ionization states depending upon the surrounding media. The use of the term, such as, for example gluconic acid, to designate such molecules is intended to include all ionization states of the organic molecule referred to. Thus, for example, both "D-gluconic acid" and "D-gluconate" refer to the same organic moiety and are not intended to specify particular ionization states or chemical forms.

As used herein the term "comprising" and its cognates are used in their inclusive sense; that is, equivalent to the term "including" and its corresponding cognates. "A" "an" and "the" include plural references unless the context clearly dictate otherwise.

B. Embodiments.

30 Polypeptides and polynucleotides:

Provided herein are isolated novel proteins, multimeric oxidoreductase complexes and nucleic acids. Also provided herein are the use of said isolated proteins, multimeric complexes and nucleic acids. Further provided are methods for the production of said proteins, multimeric complexes and nucleic acids in host cells.

Moreover, in one aspect of the invention, the multimeric complexes have been identified

- 16 -

as belonging within the family of oxidoreductases and include three subunits which are designated a cytochrome C subunit, a dehydrogenase subunit and a gamma subunit, said gamma subunit having an unidentified function.

CytC possesses a wide range of properties and functions in a large number of different redox reactions. (Pettigrew et al., CYTOCHROMES C. BIOLOGICAL ASPECTS, Springer Verlag, Berlin, Heidelberg, New York (1987). While not meant to limit the invention in any manner, it is believed that one role of CytC is to provide substrate recognition specificity. Some properties of several CytCs are described herein:

- class 1, having 1 - 2 heme groups with a redox potential of 0 to +500 mV;
- class IIa, having 1 heme group with a redox potential of 0 to +150mV;
- class IIb with 1 heme group having a redox potential of 0 to +500mV;
- class III, having 3 to 16 heme groups with a redox potential of -400 to 0mV;
- class IV having 4 heme groups with a redox potential of -80 to +400mV;
- class c₁ having 1 heme group with a redox potential of +200 to +290mV; and
- class c554 with various subgroups having between 1 and 2 heme groups with a redox potential of -276 to +47mV.

Further proteins belonging to the CytC family include the following heme consensus sequence Cys-X-X-Cys-His wherein the His residue is one of the two axial ligands of the heme iron and X represents any amino acid residue (Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol.* 45:1 - 56). Reference is also made to the PROSITE database of protein families and domains. In a preferred embodiment of the invention the CytCs have between 1 and 4 heme groups, that is at least 1, at least 2, at least 3, and at least 4 heme groups. In one preferred embodiment the CytCs of the invention have 3 heme groups. Reference is made to figure 16A and 16B wherein 3 heme consensus sequences are illustrated for each of the CytC orfs.

In one embodiment, an isolated CytC will comprise any one of the amino acid sequences illustrated in Figures 4A-4B and 5A-5B (SEQ ID NOs. 20 - 38) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity to the amino acid sequences thereof.

In one preferred embodiment an isolated CytC comprises the amino acid sequence of any one of SEQ ID NOs. 20 - 29 (Figures 4A-4B) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In one preferred embodiment an isolated CytC will comprise the amino acid sequence of SEQ ID NO. 22 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In a second preferred embodiment an isolated CytC will

- 17 -

comprise the amino acid sequence of SEQ ID NO. 23 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In a third preferred embodiment an isolated CytC will comprise the amino acid sequence of SEQ ID NO. 24 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In a fourth preferred embodiment an isolated CytC will comprise the amino acid sequence of SEQ ID NO. 29 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In another preferred embodiment an isolated CytC comprises the amino acid sequence of any one of SEQ ID NOs. 30 – 38 (Figures 5A-5B) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. More preferably an isolated CytC will comprise the amino acid sequence of SEQ ID NO. 38 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

A further embodiment of the invention includes polynucleotides encoding a CytC according to the invention. Preferred polynucleotides include those having any one of the nucleic acid sequences illustrated in Figures 11A - 11H and 12A - 12H (SEQ ID NOs. 78 – 96) and nucleic acid sequences having at least 85%, 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In one embodiment the nucleic acid sequences which encode a CytC are any one of the sequences illustrated in SEQ ID NOs. 78 – 87 and nucleic acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Preferably the nucleic acid sequence encoding a CytC is the sequence illustrated in SEQ ID NO. 80 and sequences having at least 90%, 93%, 95%, 97%, 98% and 99% identity thereto. Preferably the nucleic acid sequence encoding a CytC is the sequence illustrated in SEQ ID NO. 81 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

Preferably the nucleic acid sequence encoding a CytC is the sequence illustrated in SEQ ID NO. 82 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Also preferably the nucleic acid sequence encoding a CytC is the sequence illustrated in SEQ ID NO. 87 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In another embodiment the nucleic acid sequences which encode a CytC are any one of the sequences illustrated in SEQ ID NOs. 88 – 96 and nucleic acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Preferably the

- 18 -

nucleic acid sequence encoding a CytC is the sequence illustrated in SEQ ID NO. 96 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In a further embodiment the invention includes a polynucleotide encoding a CytC having an amino acid sequence of any one of SEQ ID NOs. 20 - 38 or an amino acid sequence having at least 95%, 96%, 96%, 98% and 99% sequence identity thereto.

The CytC sequences having said identity can be identified by substantial nucleic acid and/or amino acid sequence homology to the CytC sequences outlined herein. Such homology can be based upon the overall nucleic acid or amino acid sequence, and is generally determined as outlined below, using either homology programs or hybridization conditions. Homology as used herein is in reference to sequence identity. Homology will be determined using standard techniques known in the art, including, but not limited to, the local homology algorithm of Smith & Waterman, (1981) *Adv. Appl. Math.* 2:482, by the homology alignment algorithm of Needleman & Wunsch, (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson & Lipman, (1988) *PNAS USA* 85:2444 -2448 by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), and the Best Fit sequence program described by Devereux, et al. (1984), *Nucl. Acid Res.* 12:387-395, preferably using the default settings, or by inspection. Also reference is made to the multiple sequence alignment program CLUSTRAL W (Thompson et al. (1994) *Nucleic Acid Research* 22: 4673 - 4680).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, (1987) *J. Mol. Evol.* 35:351-360; the method is similar to that described by Higgins & Sharp (1989) *CABIOS* 5:151-153. Useful PILEUP parameters include a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

In a preferred embodiment, multiple sequence analysis is done using the Lasergene program suite from DNASTAR. DNASTAR uses the Clustal algorithm in the Megalign program version 3.12. Default multiple alignment parameters include a gap penalty of 10 and a gap length penalty of 10. Pairwise alignment default parameters include Ktuple of 1, a gap penalty of 3; a window of 5 and diagonals saved of 5.

Another example of a useful algorithm is the BLAST algorithm, described in Altschul, et al. (1990), *J. Mol. Biol.* 215, 403-410 and Karlin, et al. (1993), *PNAS USA*

- 19 -

90:5873-5787. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul, et al., *Methods in Enzymology* 266: 460-480 (1996); WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap
5 fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity. A % amino acid sequence identity value is determined
10 by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).

An updated BLAST algorithm, is described in Altschul, et al. (1997), *Nucleic
15 Acid Res.* 25, 3389-3402. A particularly useful BLAST program is Basic BLAST. Preferred parameters are Lambda K H 0.318, 0.135, 0.401 and gapped Lambda K H 0.27, 0.0470, 0.23, Matrix: BLOSUM62, gap penalties: existence 11, extension 1.

In one embodiment, nucleic acid encoding a CytC having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity with a nucleic acid illustrated in any one of SEQ
20 ID NOs. 78 – 96 is determined through hybridization studies. One embodiment involves nucleic acids which hybridize under high stringency conditions to the nucleic acid sequences identified in figures 11 or 12 or a complement thereof.

High stringency conditions are known in the art and see for example, Maniatis, et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2d Edition (1989) and *SHORT
25 PROTOCOLS IN MOLECULAR BIOLOGY*, ed Ausubel et al. Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures.

An extensive guide to the hybridization of nucleic acids is found in Tijssen, *TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY – HYBRIDIZATION WITH
30 NUCLEIC ACID PROBES*, Overview of Principles of Hybridization and the Strategy of Nucleic Acid Assays (1993). Generally stringent conditions are selected to be about 5 to 10 degrees lower than the thermal melting point T_m for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes
35 complementary to the target, hybridize to the target sequence at equilibrium.

- 20 -

Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least 30°C for short probes (e.g. for 10 to 50 nucleotides) and at 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.

In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, *supra*, and Tijssen, *supra*.

Polymerase chain reaction PCR may also be used to screen for homologous sequences and reference is made to Chen et al., (1995) *Biotechniques* 18(4):609 - 612. Other methods include protein bioassay or immunoassay techniques which include membrane-based, solution-based or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

The invention also provides a method of using the polynucleotides which encode a CytC or dehydrogenase subunit as defined herein as probes for detecting other CytC or dehydrogenase proteins in microbial organisms. In one embodiment at least 10, 15, 20, 15, 30, 40, 50 or more contiguous bases from any one of SEQ ID NOs. 59 - 77 (for a dehydrogenase protein) or from any one of SEQ ID NOs. 78 - 96 (for a CytC protein) may be used as a probe. Further oligonucleotide probes useful in the present invention may comprise a nucleic acid sequence encoding a polypeptide having at least 5, 10, 15, 20, 25, 30 or more contiguous amino acid residues of any one of SEQ ID NOs. 1 - 19 (for a dehydrogenase protein) and any one of SEQ ID NOs. 20 - 38 (for a CytC protein). Hybridization studies as indicated above would be further used to identify CytC and dehydrogenase proteins in various microbial organisms and specifically in *Pantoea* species.

In another embodiment, the invention is directed to an isolated dehydrogenase comprising any one of the amino acid sequences illustrated in Figures 2A-2C and 3A-3B (SEQ ID NOs. 1 - 19) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity to the amino acid sequences thereof. In one embodiment the dehydrogenase is a flavin-linked dehydrogenase and in another embodiment the dehydrogenase is a pyridine-linked dehydrogenase. In a further embodiment the dehydrogenase exhibits polyol dehydrogenase (PDH) activity and reference is made to Nordling et al. (2002) *Eur. J. Biochem.* 269:4267 - 4276. The PDH family of

- 21 -

dehydrogenases includes sorbitol dehydrogenases (Saito et al., (1997) *Appl. Environ. Microbiol.* 63:454 - 460).

In another embodiment an isolated dehydrogenase comprises the amino acid sequence of any one of SEQ ID NOs. 1 – 10 (Figures 2A - 2C) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Preferably an isolated dehydrogenase will comprise

- a) the amino acid sequence of SEQ ID NO. 3 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto;
- b) the amino acid sequence of SEQ ID NO. 4 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto;
- c) the amino acid sequence of SEQ ID NO. 5 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto; and
- d) the amino acid sequence of SEQ ID NO. 6 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In another preferred embodiment an isolated dehydrogenase according to the invention comprises the amino acid sequence of any one of SEQ ID NOs. 11 – 19 (Figures 3A - 3B) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Preferably an isolated dehydrogenase will comprise the amino acid sequence of SEQ ID NO. 19. In one preferred embodiment the dehydrogenase is a flavin-linked dehydrogenase.

A further embodiment of the invention includes isolated polynucleotides encoding a dehydrogenase according to the invention. Preferred polynucleotides include those having any one of the nucleic acid sequences illustrated in Figures 9A-9J and 10A-10H (SEQ ID NOs. 59 - 77) and nucleic acid sequences having at least 85%, 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

In one embodiment the nucleic acid sequences which encode a dehydrogenase are any one of the sequences illustrated in SEQ ID NOs. 59 - 68 (Figures 9A-9J) and nucleic acid sequences having at least 85%, 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Preferably the nucleic acid sequence encoding a dehydrogenase is

- a) the sequence illustrated in SEQ ID NO. 61 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto;
- b) the sequence illustrated in SEQ ID NO. 62 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto;

- 22 -

c) the sequence illustrated in SEQ ID NO. 63 and sequences having at least 90%, 93%, 95%, 97%, 96%, 98% and 99% identity thereto; and

d) the sequence illustrated in SEQ ID NO. 64 and sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto.

5 In another embodiment the nucleic acid sequences which encode a dehydrogenase according to the invention are any one of the sequences illustrated in SEQ ID NOs. 69 – 77 (Figures 10A-10H) and nucleic acid sequences having at least 85%, 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. One preferred embodiment comprises the nucleic acid sequence having the sequence
10 illustrated in SEQ ID NO. 77.

In a further embodiment the invention includes a polynucleotide encoding a dehydrogenase having an amino acid sequence of any one of SEQ ID NOs. 1 - 19 or an amino acid sequence having at least 95%, 96%, 96%, 98% and 99% sequence identity thereto.

15 Sequence identity is generally determined as outlined above for CytC sequences using either homology programs or hybridization conditions.

Dehydrogenase assays are well known and may be adopted from the methods described in Bouvet et al. (1989) *Int. J. Syst. Bacteriol.* 39:61 - 67 using cells grown on MGY supplemented with 2-KDG. Reference is also made to
20 Shinagawa and Ameyama (1982) *Meth. Enzymol.* 89:194 - 198.

Oxidoreductase complexes:

In one embodiment of the invention an isolated multimeric oxidoreductase complex includes at least two subunits and the complex is coded for by an operon
25 comprising at least two genes. The subunits include a CytC subunit and a dehydrogenase subunit as defined herein. In another embodiment the multimeric oxidoreductase complex comprises three subunits of different molecular weights which are coded for by an operon comprising at least three genes. The subunits include a CytC subunit, a dehydrogenase subunit and a third subunit which is defined by a nucleic acid
30 encoding an amino acid sequence as set forth in any one of SEQ ID NOs. 39 - 57 (Figures 6A-6B and 7) or an amino acid sequence having at least 95%, 96%, 97%, 98% and 99% identity thereto.

Preferred isolated multimeric oxidoreductase complexes include (a) a CytC subunit having an amino acid sequence of any one of SEQ ID NOs. 20 – 38 (Figures 4A-
35 4B and 5A-5B) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%,

- 23 -

98% and 99% identity to the amino acid sequences thereof and (b) a dehydrogenase subunit having an amino acid sequence of any one of SEQ ID NOs. 1 – 19 (Figures 2A-2C and 3A-3B) and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In one preferred embodiment the CytC subunit of the multimeric oxidoreductase complex will comprise an amino acid sequence selected from the group consisting of SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 29, SEQ ID NO. 38, and an amino acid sequence having at least 95%, 96%, 97%, 98% and 99% identity thereto. In another preferred embodiment the dehydrogenase subunit of the multimeric oxidoreductase complex will comprise an amino acid sequence selected from the group consisting of SEQ ID NO. 3, SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 19 and an amino acid sequence having at least 95%, 96%, 97%, 98%, 99% sequence identity thereto.

In another embodiment the isolated multimeric oxidoreductase complex further comprises a third subunit having an amino acid sequence of any one of SEQ ID NOs. 39 - 57 (Figures 6A-6B and 7) and an amino acid sequence having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In one embodiment the third subunit will comprise an amino acid sequence from any one of SEQ ID NOs. 39 - 48 and an amino acid sequence having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% sequence identity thereto. According to this embodiment, the third subunit will comprise about 220 - 265 amino acid residues, also about 225 - 255 amino acid residues and about 235 to 250 amino acid residues. In another preferred embodiment the third subunit will comprise an amino acid sequence from any one of SEQ ID NOs. 49 - 57 and an amino acid sequence having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% sequence identity thereto. According to this embodiment, the third subunit will comprise about 155 to 215 amino acid residues, also about 165 to 205 amino acid residues and about 170 to 200 amino acid residues. In a preferred embodiment, the third subunit of the multimeric oxidoreductase complex will comprise an amino acid sequence selected from the group consisting of SEQ ID NO. 39, SEQ ID NO. 40, SEQ ID NO. 41, SEQ ID NO. 42, SEQ ID NO. 57 and an amino acid sequence having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% sequence identity thereto.

In another embodiment of the invention an isolated multimeric oxidoreductase complex is encoded by a) a first polynucleotide coding for a CytC subunit having an amino acid sequence of anyone of SEQ ID NOs. 20 – 38 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity to the amino acid sequences thereof and (b) a second polynucleotide coding for a dehydrogenase subunit

- 24 -

having an amino acid sequence of any one of SEQ ID NOs. 1 – 19 and amino acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. The isolated multimeric oxidoreductase may further be encoded by a third polynucleotide coding for a third subunit having an amino acid of any one of SEQ ID NOs. 39 - 57 and amino acid sequence having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. In one preferred embodiment the first polynucleotide will encode a CytC subunit having the amino acid sequence of SEQ ID NO. 24, SEQ ID NO. 38 or an amino acid sequence having at least 98% identity thereto. In another preferred embodiment the second polynucleotide will encode a dehydrogenase subunit having the amino acid sequence of SEQ ID NO. 6, SEQ ID NO. 19 or an amino acid sequence having at least 98% sequence identity thereto.

In yet another embodiment the isolated multimeric oxidoreductase complex will be encoded by a) a first polynucleotide encoding a CytC subunit, said polynucleotide having a nucleic acid sequence set forth in any one of SEQ ID NOs. 28 - 96 and nucleic acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto and b) a second polynucleotide encoding a dehydrogenase subunit, said polynucleotide having a nucleic acid sequence set forth in any one of SEQ ID NOs. 59 - 77 and nucleic acid sequences having at least 90%, 93%, 95%, 96%, 97%, 98% and 99% identity thereto. Further the multimeric oxidoreductase complex will be encoded by c) a third polynucleotide encoding a third subunit, said third polynucleotide having a nucleic acid sequence set forth in any one of SEQ ID NOs. 97 - 115 (Figures 13A -13E and 14A - 14C) and nucleic acid sequence having at least 90% 93%, 95%, 96%, 97%, 98% and 99% sequence identity thereto.

In one embodiment the CytC subunit represents the last gene in the multimeric oxidoreductase complex operon. In another embodiment, the dehydrogenase subunit is represented by the subunit having the greatest molecular weight in the multimeric oxidoreductase complex.

Other preferred isolated multimeric oxidoreductase complexes comprise any one of the 19 operons identified in Figures 1A and B. More specifically,

- (1) SEQ ID NOs: 46, 1 and 20;
- (2) SEQ ID NOs: 48, 8 and 27;
- (3) SEQ ID NOs: 41, 5 and 23;
- (4) SEQ ID NOs: 44, 7 and 28;
- (5) SEQ ID NOs: 42, 4 and 22;

- 25 -

- (6) SEQ ID NOs: 45, 9 and 26;
- (7) SEQ ID NOs: 43, 2 and 21;
- (8) SEQ ID NOs: 47, 10 and 25;
- (9) SEQ ID NOs: 39, 6 and 24;
- 5 (10) SEQ ID NOs: 40, 3 and 29;
- (11) SEQ ID NOs: 52, 18 and 34;
- (12) SEQ ID NOs: 55, 14 and 35;
- (13) SEQ ID NOs: 53, 31 and 16;
- (14) SEQ ID NOs: 57, 19 and 38;
- 10 (15) SEQ ID NOs: 12, 50 and 32;
- (16) SEQ ID NOs: 54, 13 and 37;
- (17) SEQ ID NOs: 58, 56, 15 and 36;
- (18) SEQ ID NOs: 51, 17 and 33;
- (19) SEQ ID NOs: 30, 49 and 11; and
- 15 (20) operons having subunits with at least 90%, 93%, 95% 96%, 97%, 98%,
and 99% sequence identity to the subunits of operons (1) – (19).

In one preferred embodiment the multimeric oxidoreductase complex has 2-KDGDH activity and comprises operon (14) Figure 1, which includes the subunits having the amino acid sequences of SEQ ID NO. 57, SEQ ID NO. 19 and SEQ ID NO. 38 and
20 sequences having at least 95%, 96%, 97%, 98% and 99% identity thereto.

In another preferred embodiment the multimeric oxidoreductase complex has GADH activity and comprises an operon selected from the group consisting of

- a) operon (3) which includes the subunits having an amino acid sequence of SEQ ID NOs: 41, 5 and 23;
- 25 b) operon (5) which includes the subunits having an amino acid sequence of SEQ ID NOs: 42, 4 and 22;
- c) operon (9) which includes the subunits having an amino acid sequence of SEQ ID NOs: 39, 6 and 24;
- d) operon (10) which includes the subunits having an amino acid sequence of
30 SEQ ID NOs: 40, 3 and 29; and
- e) operons having subunits with at least 95%, 96%, 97%, 98%, and 99% sequence identity to each of the subunits of operons (3), (5), (9), or (10).

In yet another preferred embodiment of the multimeric oxidoreductase complex,
35 the subunits comprising the operons of operon (3), operon (5), operon (9) and operon

- 26 -

(10) may be mixed to form new operons. For example, SEQ ID NO. 6 which represents the dehydrogenase subunit of operon (9) could be combined with a) SEQ ID NO. 23, which represents the CytC subunit of operon (3); b) SEQ ID NO. 22, which represents the CytC subunit of operon (5) or c) SEQ ID NO. 29, which represents the CytC subunit of operon (10).

The isolated multimeric oxidoreductase complexes described above are preferably encoded by polynucleotides having the nucleic acid sequences set forth as follows:

- (1) SEQ ID NOs: 104, 59 and 78;
- (2) SEQ ID NOs: 106, 66 and 85;
- (3) SEQ ID NOs: 99, 63 and 81;
- (4) SEQ ID NOs: 102, 95 and 86;
- (5) SEQ ID NOs: 100, 62 and 80;
- (6) SEQ ID NOs: 103, 67 and 84;
- (7) SEQ ID NOs: 101, 60 and 79;
- (8) SEQ ID NOs: 105, 68 and 83;
- (9) SEQ ID NOs: 97, 64 and 82;
- (10) SEQ ID NOs: 98, 61 and 87;
- (11) SEQ ID NOs: 110, 76 and 92;
- (12) SEQ ID NOs: 113, 72 and 93;
- (13) SEQ ID NOs: 111, 89 and 74;
- (14) SEQ ID NOs: 115, 77 and 96;
- (15) SEQ ID NOs: 70, 108 and 90;
- (16) SEQ ID NOs: 112, 71 and 95;
- (17) SEQ ID NOs: 114, 73 and 94;
- (18) SEQ ID NOs: 109, 75 and 91;
- (19) SEQ ID NOs: 88, 107 and 69;

and sequences having at least 90%, 93%, 95%, 96%, 97%, 98%, and 99% sequence identity to the sequences of (1) - (19).

Preferred multimeric oxidoreductase complexes include complexes having identical or similar activity to known GADHs (Matsushita, et al. (1982) *METHODS IN ENZYMOLOGY* 89: 187 - 193; Yum et al. *J. of Bacteriol.* (1997) 179:6566 - 6572; Matsushita et al. (1979), *J. Biochem.* 85:1173; and Kulbe et al. (1987), *Ann. N.Y. Acad. Sci* 506:552) and to known 2-KGDHs (Shinagawa et al. (1982) *Methods in Enzymology*

- 27 -

89: 194 - 198; Pujol et al. (2000) *J. Bacteriol.* 182:2230 - 2237 and Stroshane (1977) *Biotechnol. BioEng* 19:459).

The various orfs of the operons as listed in Figures 1A and 1B may be subject to *in vitro* recombination, wherein fragments of the genes provide a means of generating new oxidoreductase proteins having improved enzyme activity, stability or altered substrate specificity, for example. *In vitro* recombination may be accomplished by various well known means such as by DNA shuffling and variations of this technique (Stemmer W.P. (1994) *PNAS USA* 91:10747-10751; Stemmer W.P. (1994) *Nature* 370:389 - 391; Zhao et al., (1999) *Protein Eng.* 12: 47- 53; Zhao et al., (1998) *Nucleic Acids Res.* 26: 681-683 and Arnold et al. (1997) *Adv. Biochem. Eng. Biotechnol* 58: 1 - 14.). DNA shuffling relies on homologous recombination during PCR reassembly of gene fragments from multiple parent sequences. Sequence crossovers are generated at points of high sequence identity.

Other *in vitro* recombination techniques include those described in Wells et al., (1985) *Gene* 34: 315 - 323 and Wang and Malcom (1999) *Biotechniques* 26: 680 - 682. Those skilled in the art are also aware of non-homologous random recombination to generate new gene variants and reference is made to Bittker et al., (2002) *Nat. Biotech.* 20:1024 - 1029).

Vectors:

Further the invention pertains to host cells transformed with a vector comprising a polynucleotide encoding a CytC subunit, a dehydrogenase subunit or a multimeric oxidoreductase complex as described herein. Plasmids which can be used as vectors in bacterial organisms are well known and reference is made to Maniatis et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed. (1989); Sambrook et al. *MOLECULAR CLONING: A LABORATORY MANUAL* 2nd Ed. (1989) and Bron, S. Chapter 3, Plasmids in *MOLECULAR BIOLOGY METHODS FOR BACILLUS*, Eds. Harwood and Cutting (1990) John Wiley & Sons Ltd.

One preferred plasmid for the introduction of polynucleotides encoding non-naturally occurring proteins into a strain of *Enterobacteriaceae* is RSF1010, a mobilizable, but not self transmissible plasmid which has the capacity to replicate in a broad range of bacterial hosts, including gram negative and gram positive bacteria (Frey et al., (1989), *The Molecular Biology of IncQ Plasmids*, In: Thomas (Ed.) *PROMISCUOUS PLASMIDS OF GRAM NEGATIVE BACTERIA*, Academic Press, London pp.79 - 94 and Frey et al. (1992) *Gene* 113:101 - 106).

- 28 -

Promoters useful with the polynucleotides encoding the CytC subunit, the dehydrogenase subunit, a multimeric oxidoreductase complex or any of the operons disclosed herein include the native promoter of the operon. However, both synthetic promoters and hybrid promoters may also be useful. One example of a useful hybrid promoter is a *tac* promoter, which is a hybrid of the *trp* and *lac* promoters. Promoters are well known in the art and reference is made to Sommer et al. (2000) *Microbiol.* 146:2643 - 2653; Brosius et al., (1985) *J. Biol. Chem.* 260:3539; Russell and Bennett (1982) *Gene* 20: 231; Mulligan et al., (1985) *J. Biol. Chem.* 260: 3529; Deuschle et al., (1986) *EMBO J.* 5:2987 - 2994; Amann et al. (1983) *Gene* 5: 167 - 178 and Amore et al. (1989) *Appl. Microbiol. Biotech.* 30:351 - 357. In addition to a functioning promoter sequence, an expression vector may also include an efficient ribosome binding site and a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

Gene transfer:

Gene transfer techniques for bacterial cells are well known and these techniques include transformation, transduction, conjugation and protoplast fusion. Gene transfer is the process of transferring a polynucleotide, gene or genes to a cell wherein exogenously added DNA is taken up by a bacterium. General transformation procedures are taught in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Vol. 1 Ausubel et al. Eds., John Wiley and Sons, Inc. 1987, Chapter 9). These techniques include calcium chloride methods, transformation using DEAE-Dextran and electroporation. Also reference is made to USP 5,032,514; Potter, H. (1988) *Anal. Biochem.* 174:361 - 373; Sambrook, *supra* and Ferrari et al., Genetics pgs 57 - 72 in BACILLUS, Harwood et al., Eds. Plenum Publishing Corp. Transformation of a host cell can be detected by the presence/absence of selective marker gene expression.

Host cells:

Bacterial cells are the preferred host cells according to the invention and particularly preferred bacterial host cells are Enterobacteriaceae cells. More specifically the cells of *Erwinia*, *Enterobacter*, *Gluconobacter*, *Klebsiella*, *Escherichia* and *Pantoea* and most preferably *Pantoea* sp. *Klebsiella* sp. and *E. coli*

- 29 -

cells. Particularly preferred *Pantoea* cells are *P. citrea* and *P. agglomerans*, and reference is made to USP 5,032,514; Truesdell et al., (1991) *J. Bacteriol.* 173:6651 - 6656 and ATCC accession number 39140. Other preferred bacteria host strains include strains of Pseudomonadaceae. *Bacillus* strains may also serve as host cells.

5 In one embodiment a host cell is transformed with a vector comprising any one of the 19 operons illustrated in Figures 1A and 1B. Preferably a host cell may be transformed with a polynucleotide encoding the sequences set forth in

a) SEQ ID NO. 3 and SEQ ID NO. 29;

b) SEQ ID NO. 4 and SEQ ID NO. 22;

10 c) SEQ ID NO. 5 and SEQ ID NO. 23;

d) SEQ ID NO. 6 and SEQ ID NO. 24 and

e) sequences having at least 95%, 96%, 97%, 98% and 99% sequence identity to the sequences of a) - d).

Also a host cell may be transformed with a polynucleotide encoding the
15 sequences set forth in SEQ ID NO. 19 and SEQ ID NO. 38 or sequences having at least 95%, 96%, 97%, 98% and 99% sequence identity to either SEQ ID NO. 19 or 38.

In another embodiment a bacterial host cell may be transformed with an isolated polynucleotide which codes for a multimeric oxidoreductase complex having
20 an amino acid sequence comprising an alpha subunit, a gamma subunit and a cytochrome C subunit, wherein the alpha subunit has dehydrogenase activity and is selected from the sequences of SEQ ID NOs: 1 - 19 and sequences having at least 95%, 96%, 97%, 98% and 99% sequence identity thereto; the gamma subunit is selected from the sequences of SEQ ID NOs: 39 - 57 and sequences having at
25 least 95%, 96%, 97%, 98% and 99% sequence identity thereto; and the cytochrome C subunit is selected from the group of sequences of SEQ ID NOs: 20 - 38 and sequences having at least 95%, 96%, 97%, 98% and 99% sequence identity thereto.

As stated above, bacterial host cells may be recombinant host cells.

30 Modification to host cells may have been realized prior to, simultaneously with, or after introduction of a polynucleotide encoding a CytC, a dehydrogenase or a multimeric oxidoreductase complex as defined according to the invention.

Recombinant host cells may include chromosomal inactivations, such as deletions or interruptions of endogenous chromosomal genes, modifications resulting in

- 30 -

increased expression of endogenous chromosomal genes, and inclusion of heterologous genes.

In another embodiment, the host cells may be engineered to include genes encoding enzymes known to effect the conversion of glucose or other ordinary metabolites to 2-KDG or 2-KLG. Non-limiting examples of the enzymes effecting the conversion of an ordinary metabolite to 2-KDG or 2-KLG are D-glucose dehydrogenase (Adachi, O. et al., (1980) *Agric. Biol. Chem.*, 44:301-308; Ameyama, M. et al., (1981) *Agric. Biol. Chem.* 45:851-861; Smith et al. (1989) *Biochem. J.* 261:973; and Neijssel et al., (1989) *Antonie Van Leeuwenhoek* 56(1):51-61); D-gluconate dehydrogenase (McIntire, W. et al., (1985) *Biochem. J.*, 231:651-654; Shinagawa, E. et al., (1976) *Agric. Biol. Chem.* 40:475-483; Shinagawa, E. et al., (1978) *Agric. Biol. Chem.* 42:1055-1057; and Matsushita et al. (1979), *J. Biochem.* 85:1173); 5-keto-D-gluconate dehydrogenase (Shinagawa, E. et al., (1981) *Agric. Biol. Chem.*, 45:1079-1085 and Stroshane (1977) *Biotechnol. BioEng.* 19(4) 459); and 2,5-diketo-D-gluconic acid reductase (USP Nos: 5,795,761; 5,376,544; 5,583,025; 4,757,012; 4,758,514; 5,008,193; 5,004,690; and 5,032,514).

Recovery and detection of redox products:

Methods useful for the detection of oxidoreduction reaction products include the use of high-performance liquid chromatography (HPLC) using anion exchange (*J. Chrom.* 1980, 196:163); electro-redox procedures (Pachia, 1976, *Anal. Chem.* 48:364); thin-layer chromatography; column chromatography and mass spectrometry. The skilled artisan will be well aware of appropriate controls to be applied in utilizing these detection methods. Specifically methods for detection of AsA intermediates and AsA stereoisomerisms include the use of redox-titration with 2,6 dichloroindophenol (Burton et al. 1979, *J. Assoc. Pub. Analysts* 17:105) or other suitable reagents.

Once produced, the products of oxidoreductase reactions and specifically keto-polyol derivatives such as AsA intermediates can be recovered and/or purified by any means known to those of skill in the art, including, lyophilization, crystallization, spray-drying, and electrodialysis.

Cell cultures and fermentations:

Methods suitable for the maintenance and growth of bacterial cells are well known and reference is made to the MANUAL OF METHODS OF GENERAL BACTERIOLOGY, Eds. P. Gerhardt et al., American Society for Microbiology,

- 31 -

Washington DC (1981) and T.D. Brock in BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY, 2nd ed. (1989) Sinauer Associates, Sunderland, MA.

Cell Precultures -Typically cell cultures are grown at 25 to 32°C, and preferably about 28 or 29°C in appropriate media. Exemplary growth media useful in the present invention are common commercially prepared media such as but not limited to Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. These may be obtained from for example, GIBCO/BRL (Gaithersburg, MD). Other defined or synthetic growth media may be used and the appropriate medium for growth of the particular bacterial microorganism will be known by one skilled in the art of microbiology or fermentation science. Suitable pH ranges preferred for the fermentation are between pH 5 to pH 8. Preferred ranges for seed flasks are pH 7 to pH 7.5 and preferred ranges for the reactor vessels are pH 5 to pH 6. It will be appreciated by one of skill in the art of fermentation microbiology that a number of factors affecting the fermentation processes may have to be optimized and controlled in order to maximize the ascorbic acid intermediate production. Many of these factors such as pH, carbon source concentration, and dissolved oxygen levels may affect enzymatic processes depending on the cell types used for ascorbic acid intermediate production.

The production of various desired products, such as AsA intermediates can proceed in a fermentative environment, that is, in an *in vivo* environment, or a non-fermentative environment, that is, in an *in vitro* environment; or combined *in vivo/in vitro* environments. The fermentation or bioreactor may be performed in a batch process or in a continuous process.

In vivo biocatalytic environment.

Biocatalysis begins with culturing a bacterial host cell according to the invention in an environment with a suitable carbon source ordinarily used by *Enterobacteriaceae* or other bacterial strains. Suitable carbon sources include 6 carbon sugars, for example, glucose, or a 6 carbon sugar acid, or combinations of 6 carbon sugars and/or 6 carbon sugar acids. Other carbon sources include, but are not limited to galactose, lactose, fructose, or the enzymatic derivatives of such.

In addition, fermentation media must contain suitable carbon substrates which will include but are not limited to monosaccharides such as glucose, oligosaccharides such as lactose or sucrose, polysaccharides such as starch or cellulose and unpurified mixtures from a renewable feedstocks such as cheese

- 32 -

whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. While it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon containing substrates and will only be limited by the choice of organism, the preferred carbon substrates include glucose, fructose and sucrose and mixtures thereof. Fermentation media must also contain suitable minerals, salts, vitamins, cofactors and buffers suitable for the growth or the cultures and promotion of the enzymatic pathway necessary for ascorbic acid intermediate production.

10 Batch and continuous fermentations:

The present invention may employ a batch fermentation process, a modified batch fermentation process, called Fed-batch or a continuous fermentation process. A classical batch fermentation is a closed system where the composition of the media is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. At the beginning of the fermentation the media is inoculated with the desired bacterial organism or organisms and fermentation is permitted to occur adding nothing to the system. Typically, however, a "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 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 desired product or intermediate.

A variation on the standard batch system is the Fed-Batch system. Fed-Batch fermentation processes are also suitable in the present invention 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 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

- 33 -

T.D. Brock in BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY, Second Edition (1989) Sinauer Associates, Inc. Sunderland, MA.

In one embodiment, the concentration of a carbon substrate in the feed solution is from about 55% to about 75% on a weight/weight basis. In other
5 embodiments, the concentration is from about 60 to about 70% on a weight/weight basis.

Continuous fermentation is an open system where a defined fermentation media is added continuously to a bioreactor and simultaneously an equal amount of conditioned media is removed for processing. Continuous fermentation generally
10 maintains the cultures at a constant high density where cells are primarily in log phase growth. 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
15 systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous
20 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*.

In one embodiment, the invention concerns a method of enhancing the enzymatic conversion of a polyol or carbon substrate in a bacterial host cell which comprises
25 contacting a bacterial cell which has been altered to include a polynucleotide encoding a multimeric oxidoreductase complex or a polynucleotide encoding a dehydrogenase subunit or a cytochrome c subunit encompassed by the invention with a polyol or carbon substrate under suitable culture conditions, wherein the polynucleotide is expressed in the altered bacterial host cell and allowing the enzymatic conversion of the polyol or
30 carbon substrate to a desired product wherein, the enzymatic conversion of said polyol or carbon substrate is enhanced compared to the enzymatic conversion of the polyol or carbon substrate from a corresponding unaltered bacterial host cell exposed to substantially the same culture conditions. In some preferred embodiments, the enzymatic conversion of the polyol or carbon substrate is an oxidative conversion, such as the
35 conversion of glucose to an organic or sugar acid or a keto-derivative. Examples of sugar

- 34 -

acids include aldonic acids, uronic acids and aldonic acids. More specifically, sorbitol, gluconic acid, glucaric acid, glucuronic acid and also 2-KDG, 5-KDG and 2,5-DKG. In other embodiments, the enzymatic conversion of the polyol or carbon substrate is a reductive conversion.

5 In yet other embodiments, the invention includes a method for producing a desired product, such as a sugar acid described above comprising, contacting an altered bacterial host cell with a polyol or carbon substrate (for example, with glucose or fructose) under suitable culture conditions, wherein the altered bacterial host cell is obtained by the transformation of a bacterial host cell with a vector
10 including a polynucleotide encoding a multimeric oxidoreductase complex, a dehydrogenase, or a cytochrome C protein according to the invention and producing the desired product from an enzymatic conversion of the polyol or carbon substrate.

The manner and method of carrying out the present invention may be more fully understood by those of skill in the art by reference to the following examples,
15 which examples are not intended in any manner to limit the scope of the present invention or of the claims directed thereto. All references and patent publications referred to herein are hereby incorporated by reference.

6. EXAMPLES

20

Example 1 – Sequencing of the *Pantoea citrea* genome:

Approximately 95% of the *P. citrea* genome was sequenced by constructing a representative library using standard protocols as described in Birren B. et al., (1999) GENOME ANALYSIS: A LABORATORY MANUAL., Vol. 3 Cold Spring Harbor Laboratory Press,
25 New York. Briefly *P. citrea* genomic DNA was physically sheared by sonication to produce random fragments with an average size of 0.6 to 1.0 kilobases. The DNA fragments were blunt-ended and cloned into pUC18 (New England BioLabs, Beverly, MA). Once the quality of the library was assessed by sequencing 50 - 100 randomly picked colonies, random clones from the library were picked and amplified by PCR
30 (Dunham et al. (1999) Bacterial Cloning Systems, pages 41 - 57 in GENOME ANALYSIS: A LABORATORY MANUAL. Vol. 3 Birren B. et al., (1999) Cold Spring Harbor Laboratory Press NY) and the product of the PCR reactions were used for sequencing. Sequencing was performed using Applied Biosystems 3700 DNA sequencers (Perkin-Elmer, Foster City, CA). Each of the clones was sequenced from both ends using standard and reverse
35 pUC18 primers. An average of 400 – 500 base pairs were obtained from each template.

- 35 -

With this approach, 68,389 reads were performed and the reads totaled 34,788,222 base pairs (bp). These sequences were assembled in 224 contiguous sequences that yielded 4,439,491 bp in total.

The TBLAST program was used to translate the submitted DNA into Orfs and at least 19 operons comprising three (3) Orfs were initially annotated as sorbitol dehydrogenases based on their similarity to the sorbitol dehydrogenase sequence reported by Saito et al., (1997) *Appl. Environ. Microbiol.* 63:454 - 460.

The size of the Orf designated as the gamma subunit was used to divide the 19 operons into 2 families. Family 1 includes 10 members with a gamma subunit of about 238 to 249 amino acids and reference is made to Figure 1A. Family 2 includes 9 members with a gamma subunit of about 174 to 195 amino acids and reference is made to Figure 1B. The observation that each of the 19 operons contains orfs suggests that in the natural environment of *P. citrea*, all these genes are under pressure to stay functional. Additionally analysis of the DNA sequences upstream of the genes, showed that all the genes are preceded by appropriate ribosome binding sites (data not shown).

Example 2 – Determination that orfs 2418 – 2420 (corresponding to SEQ ID NO. 115, SEQ ID NO. 77 and SEQ ID NO. 96) encode a 2-KDGDH operon having SEQ ID NO. 57, SEQ ID NO. 19 and SEQ ID NO. 38 by inactivation of the operon.

Cloning of the 2-KDGDH operon:

Strain 139-2a/Ps- was used for cloning the 2-KDGDH operon. This strain is a derivative of strain 139-2a having ATCC accession number 39140 wherein the cryptic plasmid (pS) is removed by the methods disclosed in WO 98/59054. Reference is also made to Truesdell et al., (1991) *J. Bacteriol.* 173:6651 - 6656.

Using two PCR primers, KDGF1 and KDGR1 a 2.8-kb DNA fragment encompassing the 2-KDGDH operon from the chromosome of *P. citrea* 139-2a/Ps- strain was amplified using standard techniques (Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, Cold Spring Harbor Laboratory Press (1989) 2nd Ed.).

KDGF1 5' AGTTAGCCGCTCATTTTCCTG 3' (SEQ ID NO. 116)
KDGR1 5' AGCCGCCTGGTTTTTAC 3' (SEQ ID NO. 117)

The DNA fragment was cloned into the pZeroBlunt vector having a *lac* promoter (*lacP*) (Invitrogen, Carlsbad CA) using *E. coli* TOP10 cells as the host. This resulted in plasmid

- 36 -

pKDG2 (6.32-kb). On LA+Kan50 plates, (LB media solidified with 1.5% agar plus 50ppm kanamycin) three Kan^R transformants were obtained. When checked by digesting with appropriate restriction enzymes (*EcoRI*, *ScaI* + *SpeI*, *SaI* + *SpeI*), all three transformants were found to have inserts, and the transcriptional directions were opposite to the orientation of the *lacP*.

Construction of the knockout plasmid used to delete the 2-KDGDH operon from the *P. citrea* chromosome:

In general, the strategy used to inactivate genes by homologous recombination with a plasmid has been delineated before and reference is made to Miller et al., (1988) *J. Bacteriol.* 170:2575 – 2583. This general approach was used to inactivate the 2-KDGDH operon.

The pKDG2 plasmid obtained according to the example above, was digested with *HpaI*+*ScaI* enzymes to eliminate a 0.993-kb region from the middle to C-terminus of the alpha subunit. The plasmid was then inserted with a cat cassette (1.080-kb) flanked by two *loxP* sites (Palmeros et al., (2000) *Gene* 247:255 - 264) resulting in plasmid pKDGcat1 (6.41-kb; cat runs opposite to 2-KDGDH operon). This plasmid was verified by digestion with *NotI*, *SacI* and *XbaI* enzymes. The 1.5-kb *AatII*+*SpeI* fragment containing the ColE1 Ori region was removed from plasmid pKDGcat1, and then ligated with the 502-bp *AatII*+*SpeI* DNA fragment that contains the minimal R6K origin of replication (ori) region. The R6K ori DNA was obtained by PCR using plasmid pGP704 (Miller et al., (1988) *J. Bacteriol.* 170:2575 - 2583) as PCR substrate with primers. Thus the final knockout plasmid pKDGcatR6 (5.37-kb) was obtained. *E. coli* PIR1 strain (Invitrogen, Carlsbad, CA) was transformed using the procedure described in Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1989) 2nd Ed. In this final knockout construct, a 960-bp and a 840-bp region of homology are available at the 5'- and 3'-ends of the 2-KDGDH operon to allow homologous recombination in *P. citrea* chromosome.

Transformation into a *P. citrea* strain :

After the final knockout plasmid pKDGcatR6 (5.37 kb) was verified with *HindIII* digestion, the plasmid was electroporated (Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1989) 2nd Ed.) into *P. citrea* 139-2a/PS- (pKD46) competent cells and selected for chloramphenicol-resistant (Cm^R) transformants on LA+Cm10 plates (LA is LB plus agar plates with 10ppm Cm) using well

- 37 -

known techniques. To distinguish between single and double crossover recombination events, the CmR transformants were checked on LA+Kan3 plates for kanamycin sensitivity (KanS). Nine of the 13 CmR transformants were KanS, implying that they had undergone a double crossover recombination event that inactivated the 2-KDGDH operon. Four transformants, Nos. 4, 5, 8 and 9 were checked by PCR with both internal and external primers as described below.

PCR verification of the knockout strains:

For verifying the 2-KDGDH operon deletion, two outside primers will amplify the same size band both in the wild type containing a functional KDGDH operon and putative mutants (altered strains) wherein the 2-KDGDH operon was deleted. (Reference is made to the example above wherein 0.993-kb was exchanged with 1.08-kb cat-loxP DNA).

Thus one outside primer with one cat-gene-specific primer was used to verify the recombination junctions of the putative mutants. With cat3+KDGR2 primers, all four transformants amplified the expected 1.14-kb band as compared to the unaltered strain wherein there was no amplification. With KDGF2+cat4 primers, the transformants amplified the expected 1.17-kb band. This result revealed that the four transformants had undergone a double crossover recombination event at the KDG locus as expected thereby inactivating the operon.

KDGF2	5'	GCGTCTCTGCCATTGCGTAGTTTC	3'	(SEQ ID NO. 118)
KDGR2	5'	GGGTGCGGATCGGTGTGGTTT	3'	(SEQ ID NO. 119)
CAT3	5'	AAAGTTGGAACCTCTTACGTGCCG	3'	(SEQ ID NO. 120)
CAT4	5'	CAACAGTACTGCGATGAGTGGCAG	3'	(SEQ ID NO. 121)

Removal of the pKD46 plasmid:

Since the altered strains still contained plasmid pKD46 plasmid (Datsenko and Wanner (2000) *Proc. Natl. Acad. Sci.* 97:6640 - 6645) they were cured of the plasmid as follows. Cells were grown in liquid medium without Carbenicillin (Carb) at 30°C for 3 passages (3 days) followed by plating and isolation of single colonies. All single colony isolates lost the pKD46 plasmid when tested for Carb sensitivity on LA+Carb200 plates (Datsenko and Wanner, *supra* and Palmeros et al. (2000) *Gene* 247:255 - 264).

Furthermore, no plasmid was detected in any of the single colony isolates when plasmid DNAs were isolated using standard protocols Sambrook et al., *supra*. *Pantoea* cells that were cured of plasmid pKD46 were obtained and designated WKDG4.

- 38 -

Fermentation Experiments with *Pantoea citrea*

All reagents and materials used for the growth of bacterial cells were obtained from Diffco Laboratories (Detroit, Mich.), Aldrich Chemicals (Milwaukee, Wis.) or Sigma
5 Chemical Company (St. Louis, Mo.) unless otherwise specified.

Seed Train: Culture vials which were stored in liquid nitrogen containing the indicated strain WKDG4 were thawed in air and 0.75 mL was added to a sterile 2-L Erlenmeyer flask containing 500 mL of seed medium. Flasks were incubated at 29
10 °C and 250 rpm for 12 hours. Transfer criteria is an OD₅₅₀ greater than 2.5.

Seed flask medium - A medium composition was made according to the following: KH₂PO₄ (12.0 g/L); K₂HPO₄ (4.0 g/L); MgSO₄·7H₂O (2.0 g/L); Diffco Soytone (2.0 g/L); Sodium citrate (0.1 g/L); Fructose (5.0 g/L); (NH₄)₂SO₄ (1.0 g/L); Nicotinic acid
15 (0.02 g/L); FeCl₃·6H₂O (5 mL/L of a 0.4 g/L stock solution) and Trace salts (5 mL/L - of the following solution: 0.58 g/L ZnSO₄·7H₂O, 0.34 g/L MnSO₄·H₂O, 0.48 g/L Na₂MoO₄·2H₂O). The pH of the medium solution was adjusted to 7.0 ± 0.1 unit with 20% NaOH. Tetracycline HCl was added to a final concentration of 20 mg/L (2 mL/L of a 10 g/L stock solution). The resulting medium solution was then filter sterilized
20 with a 0.2 µ filter unit. The sterile medium was added to a previously autoclaved flask.

Production Fermentor - Additions to the reactor vessel prior to sterilization included: KH₂PO₄ (3.5 g/L); MgSO₄·7H₂O (1.0 g/L); (NH₄)₂SO₄ (0.92 g/L); Mono-sodium
25 glutamate (15.0 g/L); ZnSO₄·7H₂O (5.79 mg/L); MnSO₄·H₂O (3.44 mg/L); Na₂MoO₄·2H₂O (4.70 mg/L); FeCl₃·6H₂O (2.20 mg/L); Choline chloride (0.112 g/L) and Mazu DF-204 (0.167 g/L) an antifoaming agent.

The above constituted media was sterilized at 121°C for 45 minutes. After tank sterilization, the following additions were made to the fermentation tank:
30 Nicotinic acid (16.8 mg/L); Ca-pantothenate (3.36 mg/L); Glucose (25g/L) and Fructose (25 g/L).

The final volume after sterilization and addition of post-sterilization components was 6.0L. The prepared tank and media were inoculated with the entire contents from seed flasks prepared as described to give a volume of 6.5 L.

- 39 -

Growth conditions were at 29°C and pH 6.0. Agitation rate, back pressure and air flow were adjusted as needed to keep dissolved oxygen above zero. When the sugars initially batched into the medium were exhausted, a fed-batch process was used. The production of 2-KDG obtained after a 30 hour time course with strain WKDG4 (2-KDG conc g/L = 300) was significant compared with strain 139-2a/Ps- (2-KDG conc g/L = 0). Strain 139-2a/Ps- only makes 2-KDG transiently before it is further converted to 2,5-DKG.

Example 3 - Determination that orfs 3651 - 3653 are involved in the formation of 2-KDG.

Cloning of the GADH operon:

Two PCR primers, GADHF1 and GADHR1 were used to amplify a 3.9-kb DNA fragment containing the GADH operon from the chromosome of *P. citrea* strain 139-2a/Ps-.

GADHF1 (5'- P- CGGTACTGAGGCAATGTCATG 3'), SEQ ID NO. 122 and GADHR1 (5'- P - aCGGAGAGCCGGATATTACAT,3'), SEQ ID NO. 123 wherein P = phosphate and ATG of SEQ ID No. 123 is the start codon of the orf of the gamma subunit of GADH.

The DNA fragment was digested with restriction enzymes (HindIII, PstI and SalI) to check whether the amplification was correct. The fragment was cloned into the HindII site of the low-copy number vector pCL1920 containing a *lac* promoter (*lacP*) (Lerner et al., (1990) *Nucleic Acid Res.* 18:4631). Among 16 white colonies on LA+Spec50+X-gal+IPTG plates, (LB plates solidified with 1.5% agar plus 50ppm spectinomycin plus 40 µg/ml X-gal plus 50µM IPTG) only two colonies (Nos. 7 and 11) contained the insert. When checked with appropriate restriction enzymes (HindIII, PstI, SpeI+Clal; and XbaI + BglII), the transcriptional directions of inserts in both plasmids were the same with the plasmid borne *lacP*. One of these plasmids was named pCLE2-7 (8.45-kb) and used as an overexpression plasmid of the GADH operon and for construction of the knockout plasmid as described below.

Construction of the knockout plasmid:

The pCLE2-7 plasmid was digested with Bsu36I+Clal enzymes to remove the 910-bp region from the middle of the dehydrogenase subunit and then the cat cassette (1080-bp) flanked by two loxP sites (Palmeros et al. (2000) *Gene* 247: 255 - 264) was

- 40 -

inserted, resulting in plasmid pCLE2Cat1 (8.62-kb; cat gene runs opposite to the GADH operon). The 4065-bp GADH-cat-loxP gene cluster was amplified by PCR from the above plasmid using the GADHF1 and GADHR1 primers and ligated with the 505-bp HindII+EcoRV R6K minimal origin of replication (Miller et al. (1988) *J. Bacteriol.* 170:2575 - 2583 and reference is made to the example above) from plasmid pGP704 to generate the final knockout plasmid pE2CatR6-1 (4.58-kb). This was followed by transformation into *E. coli* PIR1 strain (Invitrogen, Carlsbad, CA) using the procedures described in Sambrook, *supra*. In this final construct, a 1160-bp and a 1825-bp region of homology are available at the 5'- and 3'- sites of the GADH operon to allow homologous recombination in the *P. citrea* chromosome.

Transformation into *P. citrea* strains and verification of the GADH inactivation:

After verifying the final knockout plasmid with appropriate restriction enzymes (AatII + AccI and BglII + SnaBI), the plasmid was transformed as described above into *P. citrea* 139-2a/Ps-, and transformants were selected on LA+Cm12 plates. Hundreds of CmR colonies appeared. Four colonies were checked for deletion of the GADH operon by PCR.

Two outside primers will amplify the same size band both in the wild type and putative mutants (see above, 910-bp coding region was exchanged with 1080-bp cat-loxP DNA). Thus, one outside primer with one cat-gene-specific primer was used to verify the putative mutants. Only No. 2 transformant amplified the expected 2.46-kb band with cat3+E2R1 primers:

Primer E2R1 5'-GCAGCCGCTACGCAGATAAAA-3' SEQ ID NO. 124

However, three other transformants (Nos. 1, 3 and 4) amplified the expected 1.64-kb band with E2F1+cat4 primers:

Primer E2F1 5'-CTCGGCGAAAAAGAACCAGACAAG-3' SEQ ID NO. 125

This result revealed that only No. 2 transformant had undergone a single crossover recombination event at the 3'-end of the GADH operon, and the other transformants (Nos. 1, 3 and 4) underwent single crossover recombination events at the 5'-end of the GADH operon. Since a single crossover recombination event will not inactivate the operon, we isolated double crossover recombinants from the No.2 transformant as described below.

The No.2 transformant was sub-cultured in liquid medium followed by isolation of single colonies. Five out of 5 single colonies isolated had undergone a second cross-over

WO 2005/012547

PCT/US2004/024204

- 41 -

at the 5' end of the GADH operon upon sub-culturing. This resulted in a true GADH inactivated strain in the 139-2a/Ps- background.

To confirm further, PCR products from the above mentioned single colonies were digested with appropriate restriction enzymes. The 1.64-kb product obtained with E2F1+cat4 primers was digested with XbaI enzyme. The 2.46-kb product obtained with cat3+E2R1 primers was digested with PstI enzyme. Both digestions showed the expected patterns for all samples. Thus, the GADH deletion strains were further confirmed by this strategy. One single colony isolate from above was named WTE2-1 and saved.

10

Fermentation studies:

Both strains, WTE2-1 (the GADH deletion strain) and 139-2a/Ps- (the wild-type control strain) are basically grown as outlined above for the 2- KDGDH deletion strain and accumulation of gluconate is measured.

CLAIMS

1. An isolated polynucleotide coding for a multimeric oxidoreductase complex, said complex having oxidoreductase activity and comprising a cytochrome C subunit comprising a sequence having at least 90% sequence identity to SEQ ID NO: 5 and a dehydrogenase subunit comprising a sequence having at least 90% sequence identity to SEQ ID NO: 23.
2. The isolated polynucleotide of claim 1, wherein said polynucleotide codes for the multimeric complex comprising a cytochrome C subunit comprising a sequence having at least 95% sequence identity to SEQ ID NO: 5 and a dehydrogenase subunit comprising a sequence having at least 95% sequence identity to SEQ ID NO: 23.
3. The isolated polynucleotide of claim 1, wherein said polynucleotide codes for the multimeric complex comprising a cytochrome C subunit comprising SEQ ID NO: 5 and a dehydrogenase subunit comprising a sequence having at least 95% sequence identity to SEQ ID NO: 23.
4. The isolated polynucleotide of any one of claims 1 to 3, wherein the polynucleotide includes a third coding region for the multimeric complex comprising a sequence having at least 90% sequence identity to SEQ ID NO: 41.
5. A multimeric oxidoreductase complex having oxidoreductase activity and comprising a cytochrome C subunit comprising a sequence having at least 90% sequence identity to SEQ ID NO: 5 and a dehydrogenase subunit comprising a sequence having at least 90% sequence identity to SEQ ID NO: 23.
6. The multimeric oxidoreductase complex of claim 5, wherein said complex comprises SEQ ID NO: 5 and SEQ ID NO: 23.

7. The multimeric oxidoreductase complex of claim 5 or 6, wherein said complex has gluconate dehydrogenase activity or 2-keto-D-gluconate dehydrogenase activity.

8. An isolated polynucleotide coding for a multimeric oxidoreductase complex which functions in the enzymatic conversion of a carbon substrate to a desired product, said complex having an amino acid sequence comprising an alpha subunit, a gamma subunit and a cytochrome C subunit, wherein

the alpha subunit has dehydrogenase activity and comprises a sequence having at least 90% sequence identity to SEQ ID NO: 5;

the gamma subunit comprises a sequence having at least 90% sequence identity to SEQ ID NO: 41; and

the cytochrome C subunit comprises a sequence having at least 90% sequence identity to SEQ ID NO: 23.

9. An isolated polynucleotide coding for a multimeric oxidoreductase complex having oxidoreductase activity and which functions in the enzymatic conversion of glucose to a desired product, said complex comprising an alpha subunit, a gamma subunit and a cytochrome C subunit, wherein

the alpha subunit comprises a sequence having at least 90% sequence identity to SEQ ID NO: 5;

the gamma subunit comprises a sequence having at least 90% sequence identity to SEQ ID NO: 41; and

the cytochrome C subunit comprises a sequence having at least 90% sequence identity to SEQ ID NO: 23.

10. The isolated polynucleotide of claim 8, wherein the desired product is an ascorbic acid intermediate.

11. A vector comprising a polynucleotide as claimed in any one of claims 1 to 4, 8, 9 or 10.

12. A bacterial host cell transformed with the vector of claim 11.
13. The bacterial host cell of claim 12, wherein the bacterial host cell is a *Enterobacteriaceae* strain.
14. The bacterial host cell of claim 13, wherein the strain is a *Pantoea*, *Klebsiella*, *Erwinia* or *E. coli* strain.
15. The bacterial host cell of claim 12, wherein the bacterial host cell is a *Pseudomonadaceae* strain.
16. A method of enhancing the enzymatic conversion of a polyol substrate in a bacterial host cell comprising:
 - contacting an altered bacterial host cell with a polyol substrate under suitable culture conditions, wherein the altered bacterial host cell is obtained by the transformation of a bacterial host cell with a vector according to claim 11, wherein said polynucleotide is expressed in the altered bacterial host cell, and
 - allowing the enzymatic conversion of the polyol substrate to a desired product, wherein the enzymatic conversion of said polyol substrate is enhanced compared to the enzymatic conversion of the polyol substrate from a corresponding unaltered bacterial host cell.
17. The method according to claim 16, wherein the enzymatic conversion of the polyol substrate is an oxidative conversion.
18. The method according to claim 16, wherein the enzymatic conversion of the polyol substrate is a reductive conversion.
19. A method for producing a desired product comprising:

contacting an altered bacterial host cell with a polyol substrate under suitable culture conditions, wherein the altered bacterial host cell is obtained by the transformation of a bacterial host cell with a vector according to claim 11 and wherein the polynucleotide encoding the multimeric oxidoreductase complex is expressed in the altered bacterial host cell, and

producing a desired product from an enzymatic conversion of the polyol substrate by the multimeric oxidoreductase complex to said desired product.

20. The method according to claim 19, wherein the polyol substrate is glucose, sorbitol or glycerol.

21. The method according to claim 19, wherein the desired product is an ascorbic acid (AsA) intermediate.

22. The method according to claim 21, wherein the AsA intermediate is 2-keto-D-gluconate or 2,5-diketo-D-gluconate.

23. The method according to claim 19, wherein the desired product is an organic acid.

24. The method according to claim 19, wherein the host cell is selected from strains of *E. coli*, *Pantaea*, *Klebsiella*, and *Erwinia*.

25. The method according to claim 19 further comprising recovering the desired product.

Figure 1A

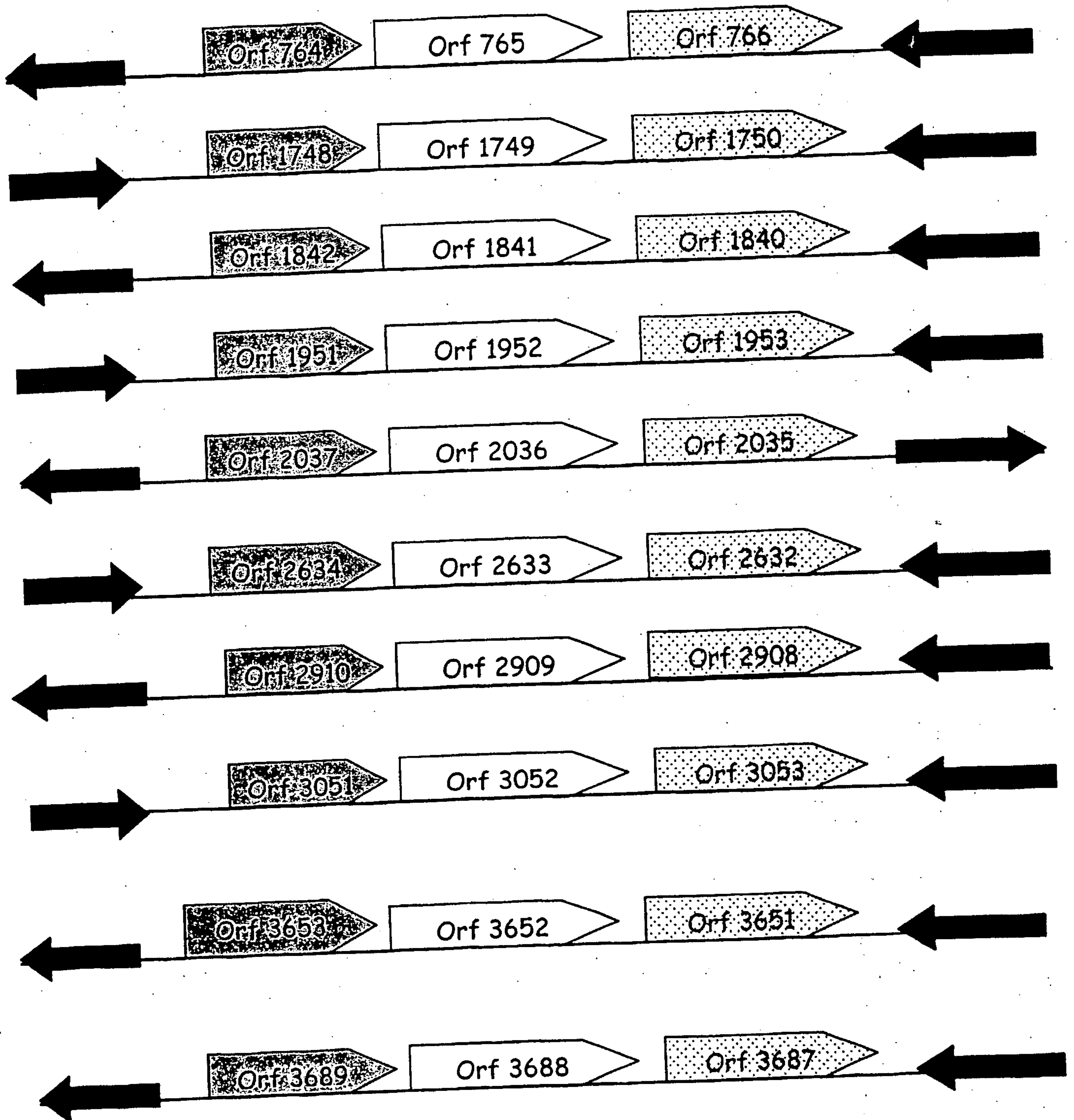
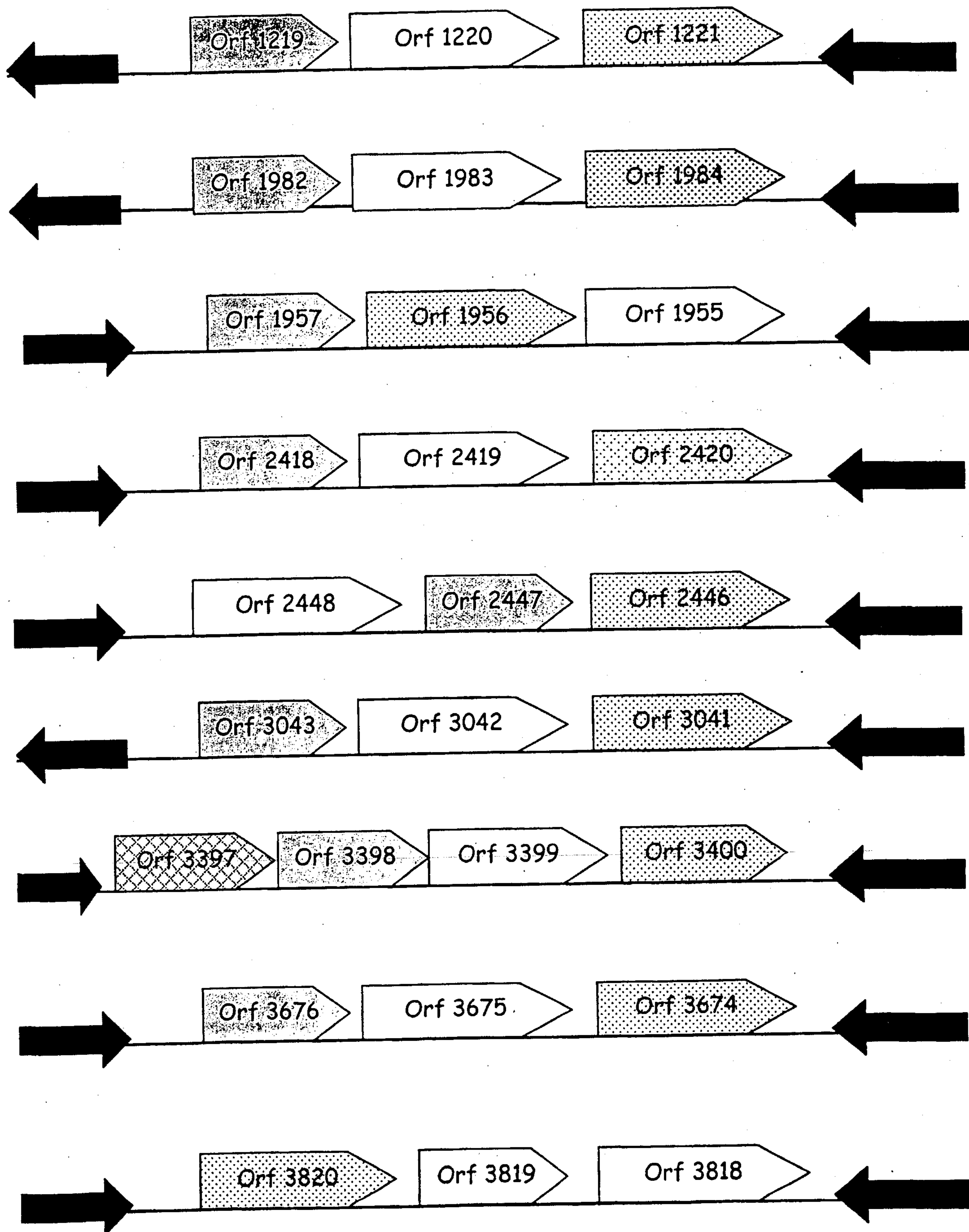


Figure 1B



WO 2005/012547

PCT/US2004/024204

FIGURE 2A

SEQ ID NO.1 - Orf 765

MAQITKKEVDVVVCGFGWAGSLMSIELAMAGLTVRALEKGPEDYEAFAYPKPADEYAYA
 VRNKVMTTPADSAVTVRYTMQDTALPTRKWGAFVPGGGVGGAGMHWTVLLRPTPTDIKL
 KTYADEAYKPGVLQEDMRVRDFPFTWNEIEPWFEKFEHICGLSGNTGNLRGQIMEGGDPF
 EGPRANPMPLPALENTLNNVMFGDTVKKMGYHPFTIPSAASRVWNTNPYGNTIAPCNYCG
 YCSKYPCLNYSKASPQTAVLDSLKQMDNFSYEVNAEVLRVVLHDDKKTAKGVIYIDEQGN
 ECFQPAKIVILSSSFQFYNVRLMLLSGIGQPYNPVTEEGVVGRNYAFLSNGSATLFFKDKN
 FNPVSSGPTGMQFNDISPGNFDGPGGLGIIGGAKIQSAQSTGTPISTALPPGTPSWGAGW
 KEGLENWYGHSMKVGITTSYRDMVYLDLDPTYKDRHGQPLLRTFNWKHNELOLQOYL
 KGIVGNIVKEMNPDSFSMSFLPMGADFDLTXYVSTHNVGGAIMGDDPKTSALNRYLQSWD
 VHNVFVPGGNAFPQNFQSNPTNTIGAITLMAANAIKEQYLKNPGPMVQV

SEQ ID NO.2 - Orf 2909

MKITNEPVDVVIVGLGWTGAIQGIELARTGLKIRALERGADRTSAEFAYPVPADLAYTK
 RHKIMQSPAAVAFTRHNLNEVALPMRELGSFRLGDGVGGADLHWTAMITRPTPVDLKL
 TYAREKFEKSQDKELRIYDFPVSWSIEIEPHMDFDQVCGSSGQAGNVRGQILPGGDPFE
 GPRSSPFPNPPLIDTLNSSMFRQAATEMGYHPYSIPSAAVSQAFNTNPYGQIAPCNYCGY
 CQFYSCNYSKASPQTALDRLKQYDNFDYKTHANVIRVEKHADGKTATGVTYIDENDNE
 VFQPAKIVILASFGLNNVRLLLNSKIGQPYNPVTEEGVVGRNYTHQYGGGITYFNQLEF
 NPFATAGPTGVVITDFGTGNINTADLGGIGGAKIYSSQPTGTPMGAPVIDSAKWGSRWK
 KGLKQSYGHSMKLEGSNMATQTNLYLDLDPNYKDKFGMPLLRVTYDYVQNDLRMLQFMR
 EKMVGIAEHLKPDHYSVGMKMDSHFASSPAYANTHAGAIMGDNPKTSVNVRYLQSWD
 VHNVFVMGACVFPQNVYANPTALVAGLTYWSAKAIRETYLNNPGPLVQA

SEQ ID NO.3 - Orf 3688

MATTKKPAADIVIVGFGWTGSLMARELADSGLKIVALERGEQRDTYPDFAYPRITDELTY
 GIRLKLQNAARETVTVRHTSSQTALPYRRFGSFLPGNGVGGAGVHWNGMLWRPLAADLK
 MHSTLVEKYGANFIPQDMTVQDYPFTYEEMEPFFDKFEKICGASGQAGNLNGEIQSGGNP
 FEQPRQNPYPTKPLQRLYAGDVFAKAAEKMGYHPFPCPAANCTEAWTNPYKVQLGVCNYC
 GFCERFGCFNYSKSGSPQSCVIPSLKAYDNFELRTNAQVIRVNTDNTGKQATGVTYIDGSG
 NEVEQPASLVILSAFQLHNVRLLLLSKIGKPYDPQTGEGVVGRNYAYQMTGGSKLFFGPD
 QDFKPFAATGTTATFIDNFNAENFDHSSLGFGVGGSTISAAFSGGRPIQQTLLPSDAPRWG
 SGWKTAIKTHYAHTMSIGASGSVMYPYQCYLDLDPTYHDVNGQPLLRTFDWQPNELKMT
 EFIGGKVEEIIKVINPPHYEMGFMMNSHYDVRPYQSTHTTGGAVMGDSPTS SVVNKYLO
 SWDVPNLFVLGACCFPQNLAYNPTGIVCATALFSAHAIKTRYLAAPGPLVTI

SEQ ID NO.4 - Orf 2036

MTIKKDPVDVVIVGFGWTGSMAMELAETGLKIVALERGEQRDTYPDFAYPRIVDELTYG
 VRLKLQNLNETVTVRHAPGDLALPYRKMGSLPGDGVGGAGVHWNGLLWRPLETDLRL
 KSTITEKYGAAFIPQDMTLQDYPFTYAEMEPFFTRFEKICGASGQAGNINGEIQGGNPF
 EAPRSGAYPTSALKSQYSGELFGKVAKEHGYSPFPGPAAICTESYQNPYGVQLGVCNYCG
 FCERFGCFNYSKASPQTCVIPALRQHTNFELRTHSHVIRVNKDSTGKKATGVTYIDANGQ
 EVEQPAALVVLGAFQLHNVRLLLLSGIGQPYDPRTGEGVVGRNYAYQVNGGVKLFYDKDQ
 YFNNFAATGCSGTIDNFNGENFDHSSLGFIGGGTISAHATGGRPIQQTTELPSGSPKWGT
 GWKKAMKDNYLHSMVGSASSVMYPYKQCYLDLDPTYTDGYGLPLLRTFDWQENDLRVTQ
 FVAGKTEELVKALKPRSYDMGFKLNTHYDVRPYQSTHTTGGAVMGDNPTS SVVNKYLO
 WDVPNVFLGACCFPQNIAYNPTGIVGATTLFAAHAIKTQYLRNPGPLVQA

WO 2005/012547

PCT/US2004/024204

FIGURE 2B

SEQ ID NO.5 - Orf 1841

MAIVKNKTDVVIIVGMGWTGAIMAKEMTDAGLSVVALERGAARDTEPDFAYPGVVDELQGS
 VHRRYLQSLHQETVTVRHNTGSVAVPYRQMGSFKPGTGVGGAGSHWSGCHFRPLPEDLRL
 RSNLEERYGKSFIPSDMTIDDFPVSYDELEPHLDMFEKVCGTSGKAGVIRGVVQAGGNPF
 EGSRSGEYPLGPNPNYLGAWEFYKAAREKGYHPYPIPASNAAGPYINPYGCQMGPNCNACG
 FCSDYGCLNYSKASPNICIMPVLRQRKNFELRTHAQVLKVNLSDDGKKATGVITYLDSNGQ
 ETEQPADLVLLCAFSLYNVHLMLLSQIGKPYDPVSNEGTVGRNYSYQNLNRVMMFYDQSV
 QANGFIGIGGSGTTMDDLNGNQLDNAQAGFVGGGI IWARQPGNGPVRGVAVPKGTPGWGS
 AWKKAUSESFRHSFYEVQACMSYQQNYLSLDPTWKDAFGRPLLRRMTFDWQPNVVKASQ
 FLVGKAVDMCQVLNPKSISSDAKKDGAHYDITKYQSTHTCGGAVMGSDPKKSALNRYLQS
 WDVPNVFAIGANAFQNNNGYNPTGLVGGGLAYWAATAIREQYLKNPGPLVQA

SEQ ID NO.6 - Orf 3652

MASVMKKTDVVIIVGFGWAGAIMAKELTEAGLNVVALERGPQRDTPDGAYPQSIDELTYN
 IRKKLFQDLSKSTVTVRHNPSTAVPYRQLNAFLPGTGTGGAGLHWSGVHFRVDPALRL
 RSHYEERYGKDFIPQGMTIQDFGVSYDELEPFDDQAEKVFGTSGTAWTVKGELVGKKGKGG
 NPFAPDRSSDFPLKAQKRTYSAQLFAEAAESVGYHPYDMPSANTSGPYTNTYGAQMGPNCN
 FCGYCSGYACMYSKASPNVNILPALRQEPKFELRND SYVLRVNLTDKKRATGVITYVDA
 SGREFEQPADLVILSAFQFHNHLMLLSGIGKPYNPVTNEGTVGRNFAYQNI STLKALFD
 KNITTNPFIGAGGAGVGVDNFADNFDHGHGFGVGGSPLVNQAGVKPISGLPTPTGTPA
 WGSEWKAADVADTYTHHVSMDAHGAHQSYRTNYLDLDPNYKDVHGQPLLRRMTFDWQENDIK
 MSQFMVSKMHNIAQAMNPKMIMGGPKTAGTHFDTTVYQTTMNGGAIMGEDPKTSAINRY
 LQSWDVSNVFVPGASAFPOGLGYNPTGMVAALTYWSAKTIREVYLKNPGPLVQA

SEQ ID NO.7 - Orf 1952

MTKKLPATDVVIIVGLGWAGSILAKELCDQGLNVIGLERGPWRDTAKDFNVATAPDELRYN
 AREELMLRPAQNTCTMRNNPSETALPMRFWGSFHPGNGTGGAGNHWAGITFRYQPADFRL
 ASHLRERYGKEVDPALTLDWGITWEEMEPFYDSFERVAGISGKAGNIKGSII EGGNPFE
 GPRARDYPNPPNIQTIAQTFFAKTATEMGYKPFNVPSALASQGYTNQYGVMTGMPCTYCGF
 CTNYGCANYSKASAI NVNLPVAVSMPNFEARTNCEVMEVLKDSSGKKATGVVYIDSNGER
 YEQPASIVIVAAFTFENVRLMLLSNVGPYPDPVTGKGTGRNYCYQTANGVRLFFKDQIF
 NPFIGGGAIGMGIDEFNNDNFDHSGLGFGVGGSTRVTPIGAAPIASRPVPPGTTPRWGSAW
 KKATVEHYLTNMSIGCEASSYPQRTNYLSLDPNYTDPHGRPLLRTFDNPNDRMAQYV
 TNKVGEIALRMNPVQIQKQPRTPAWANNQYQSSHVVGFGFVMGADPSTSAVNKFCQVWDIP
 NLFVVGGSAPVNNPGYNPTGTVGALAFRTAHYIRTQYLKQPGEMMV

SEQ ID NO.8 - Orf 1749

MSKIRPKADAVIIVGLGWAGSLMANELTQAGLNVVAIERGSWRDTSTDFPTSIDTDELRFV
 SRRAIMQPTAVETMTFRNNPLQQAALPLREFNTYQFGMNVGGAGTHWNAMTWRFLPNDFQT
 YTNTVERYGKNKFLEGMQVQDWGVITYDDLEPFYDKFERFAGTSGKAGNIKGEKIDGGNVF
 EGPRSRDYPLPLKRTQLSMIFDKATREMGLHPFAVPAGNTSGAYTNTLGINMAPCTYCG
 FCEFFGCGNWSKSSPNACILPAVMQRSNFSVITESEVLRVNKAADGKTATGVTFIGSDGV
 EWEQPADIVIIISAYQFDNVRLMLLSGIGEPYNYKTGTGVVGRNYAYQTISGAGVFFENEN
 LNPFIGAGALAAVDDYNSDNFDHSLNLDFIGGGVALVHSSNGRPIALSGAVPPGTPKWS
 KWKQAAQOSYQNYNSVYVMGNSYPHRDVFLDLDPYKDRHGQPLLRTFDWIENDKRS
 FMADRSVEIGHAMGAKTVVRQEPTARNFSPMDNLSSTTGGACMGDDPKTSAVNRYLQSW
 DVHNVFVCGASAFANNGGYNPTGTVGALTWAAEAIKNQYLKSPGPLVRI

SEQ ID NO.9 - Orf 2633

MNNIRPKADVVIIVGLGWCGSLIAEELTRAGMNVVAIERGPWWETSTDFPPSIDTDELRWD
 TRRSMLLPVAVETTTFRNNTSQQAALPSRDWNLNELGYNVGGSGTHWAGMAWRFTPFDQFQ
 YSQTvaryGKQIIVPGLILQDWGVSYDELEPFYDRFEKIAGVSGKAGKSNGNVVPEGNPF
 EGNRSSEYPLPLESTRLTDLFEQGAksLGLNPFMVPAGQASRAYVNPLGVRMGPCCTYCG

WO 2005/012547

PCT/US2004/024204

FIGURE 2C

YCLYYGCGNFSKSSPNACVIPALMQRENFTVLTDSAVVKVNKAEDGKTATGVTFIDKNNK
QWEQPADIVILSAFQMNVRLLLLSQIGQPYNPQTKQGVVGRAYSFQTVSGASLFFKDEY
LNQYIGAGALSQQVDDFNGDNFDHTGKGFIGGAGILVVARGARPIGNADTLPPGTPRWGK
EWKQAYTHAFQNTFIFGQGTSYSHEDYYLDLDPEYKDKYGLPLLRTVTFDYNDRRSK
FVEQRSVEIGKAMGAERVFGTNSASGHYSPYNFASDHTIGGAVMGTDPRTSVLNRYQQSW
DVHNVFVLGASSFPNNAGYNPTGTIGALS LWTAKAIEQYRKNPGPLVKV

SEQ ID NO.10 - Orf 3052

MNYTRPKADAVIIGLGWAGSLMAEELTRAGLNVVAIERGPWEQTQSNFSPAIAADELRYG
VRREILKPPRVETLTFRNDSSQKALPARDWNAFQMGYSVGGAGKHAANAWRFNPSDFEM
ATRHKERYNNMPLADGLILQDWGVSYAELEPFYDRVEKIAGISGKAGVLNGSTQEGGNPF
EGNRTSEYPTPPLIRSHWNTDFHNITTKMGYHPFPIPAGTIGAAFTNPLGINLAPCTYCG
YCGFYGCGNWSKSSPNICVVPALMDRTNFTLLTECTALYINKADDEKTVTGVTFRDSDGN
TGFQPADIVCLSAYQLDNVRLLLLSKIGKAYDHATGEGTLGRAYNYQTMSMGYLYYENEY
MNPFIISTGALSTQIDDFNGDNFDHTGLGFLGGAGIQALSDQGTPLSMTDRLPAGSKMWGS
AWKKAFRHSYQNYAKIQGQGTSYSHRDSYLSLDPNYTDENGQPLLRLTFDYNQNDRLMAR
FIRDRIEDICKVSGASSWITEAFPD SHNSPFRAYDSSHTIGGAVMGLDPKTSVLNRYQQH
WDAHNLFVLGASSYPNNGGYNPTITLSALT LWTAAHIVNDYLNKPGSLVR

WO 2005/012547

PCT/US2004/024204

FIGURE 3A

SEQ ID NO.11 - Orf 3818

MSQNNVDAEVIIIGSGVMGGLLATQLSAAGKSVIIVEAGPRVTRQQIVDRFRNSPFKMSL
 TNMKLQGVGSPYPDLPHVPSTYGNYLQQVGPVKYPTKYLRVVGTTWHFGSALWRMIPND
 FKLKTLYGHGRDWPFPGYDELEPWYCEAEHALGVSGVDGQDESGHGGKPWPPRSKPFPMPG
 LPTSYMFDRLSELLGKGGYNPVLEPNGRATRPWGNRPVCAGNNNCNPVCPIGAKYDGSMH
 IDQAERLGAKLLDNSVVYKIEADDNGKITRIWYKKPDGSEHSLTANLFIVAAYGIESPKL
 LLMSTSEKYPNGIANSSDQVGRNLMGHTGISMNFMAEDVWPGQGPTTELLVYLNNRDGEF
 RKTFFPSYKIKVRNTVPTADYASGLISKGVLGSELDEQLRKLSARSLNFAIDFETVPLPEN
 RVVPSKTKTDAIGIPLPEISYSVTDYWQAGKEAGLKDFANFAKLLGGDVLKIDTNYQDRQ
 HIMGTTIMGDDPKNSVVNSDCRTHDHPNLYIAGTSVMPSASCMNPTLTGAALSLRLANHL
 LKNVLV

SEQ ID NO.12 - Orf 2448

MANTTLDFDYVIVGSGVTGALIAWQLSRHGKRVCMIEAGDHIQRWKAIEHYRSLPDKSIA
 NNSPYPNLEWAPNPIGGHYLEQKGPVNYATTYIRMVGGTTWHWDSATWRLLPSDFELKTR
 YGVGRDWPIGYEVLEPWYQKAEQGLGVNGWDTEDQSGQGDHYPYPPRSQPYPTPGHPFSWG
 QQVVAGKLEAAGYSAIHEPNARLSVATAERPACAGNNTCDPICPIGAKYTADFHVQKALD
 HGCTLLSNSVVYRVEAGDDGKITAVHFRRPDKSTGTVSGKVFVIAANAIETPKLLMSVS
 ERYPOGIANTSGQVGRNLMHTGLGFNLVTEDEVWPGTGPNALLVMLNAREGKFRAERAS
 YKTKFRNTAVNFAVTKSLIKQGIMGNELYRQIKYQSARQLSIAVDLETLPNPQNRIVPSK
 DRTDSLGIPIVPEIHYDVDDYWNKGRDAAIADVQNIKILNAKIVATDTNKQNRHILGTM
 IMGNSPTDSVVDKNCRTHDHPNLYIAGTSVFPVAVGCVNPTLTGAALALRIADTLLQDPVT

SEQ ID NO.13 - Orf 3042

MKTTHSATVVIIGSGIAGSQIAQKLQKAGIDTLMLEAGSRIERWKIVENYRNSPFKTDFO
 SPYPPTRHAPHPOQYSPEDNGYFIQYGPEPYKAGYLRVAGGTTWHWSAQAWRLLPNDMRLK
 TLYGVGRDWPISYDDLEPYYYESEVEMGVGGPEDTGSPRSKPYPHPPLPLSDFDKAFKNV
 VDKNGYHLITEPAARNTPEFDGRPACCGNSNCMPICPIEAQYTGETAVRKAERAGSLLVP
 DAVVYKIEHDAKGNITSVLYKDPNGESFRVTGKIFVLAANAIETPKLMLISRSKYPNGI
 GNTTDNVGRHLMHDPGTSVYFLSKEPMWPGRGPMRLSCINNLRDGDGRSEHSAMKINLGN
 YSPTLAVSNYLLSKGVSGKDLPAVVRDYASRWAVNTFFDILPDRDNRIVAVDSQKDAMG
 IPKPGVHYHINDYINKARDVAHQHFDHIAGLFGGTEVRHDDKYFNNNHIMGTLIMGNDPN
 DSVVDADLRTHDHQNLFFVASSGVMASAGTVNCTLTLSALAMRLADKLIAECQHL

SEQ ID NO.14 - Orf 1983

MATEFDADVIVGSGACGSNLANELAVKGKSVILLEAGANVPRWKILENFRNSGRHYDRN
 NAYPNNPWSPTSNTPGYIENVGEFREQPGMLKLVGGTTWHWGGATWRYIPNDFKLKTMYG
 VGRDWPIISYSDLEPFYTRAEYAIGVAGSDTEDQSGQNPGISFPFRSKAYPVDPEADIYSN
 AKLKAALLPHGHSVVHEPTVRIHRPYDGRPGCQGNNDQVCPIGTLYNGSVHADKAVRN
 GAKLITDAVVHKITKGEQKITSVSYLTPAGEEHTLTAKYFVLAHFSFETSKLMLMNDIG
 NSSDMVGRNLMDHIGLSMFLADEPMWAGRGPVQQATIMTWRDGDGRSKYSANKHSLANN
 NPQIDIAQRAINEGLMGKELDARILDWSSRWMSIYSFLEPLPNPANRVQPNPAWKDSLGL
 PGIKVTFDVDDYTKLGAKHMEQYKQIAGLMNGQIIDLNTAFENHDHLMGMTMIMGDNPKD
 SVVNHECRSHDHPNLFIAVSGVI PAAGVVNPTLTGVALAIRSADI IAKEV

SEQ ID NO.15 - Orf 3399

MSDSLADVVIGAGIAGSLAALKMAKAGASVLILESGPEIKRDEAVNYFRNSPFGDFT
 EPYPPEPWAPQPKFIPTDNNYLIQKGPDPYRAQYLRGIGGTTWHWAGQAFRLLPNDMKIN
 TLYGVGRDWPISYEDLEPYYSDAEYQMGVSGDDDLNSPRSRPYPLPGIPLPYGFERLKQR
 LSPLGYQVGIGPQARNIPYQGRPACCGNNNCMPVCPIDAQYHGGISARKAVDAGVKIIA
 NAVVYRIEADDHGVIQAVHYLDQNKATHRVTKQFVLTAANGVESPKILLSTSDRYPNGI
 ANSSGMVGRNLMHDPGTSVEFYADEPIWFGRGPMRPGSINNMRDGSWRSESRALRIDLAN

FIGURE 3B

TSPVRYLTERLVVRQGYGKALNDKLAFAQERFVQLKCLLEMLPDPENRLVLSKTEKDAWG
IPRLEVYYKFPPEYVHAGYDQMSDFRKIVQQMGTEPLYSQRGVYDNNQHITGTMIMGSD
PKNSVVDGNCRTHDHPNLFIAGTGIMPSASTVNSTLTGTALALRMADYVLKSL

SEQ ID NO.16 - Orf 1955

MSEQYSADV VVG GICGGTVAKELAEAGLSVLVLDAGPRWERGEVVENWRNLPPV NKSE
SDYATPYPAPPAEPWAVHPQLYPYNNYPEVSGPDASAFRQGMKGVGGTTWHWAASCWRFLPA
DMQLQTTYGVGRDWVVTYDEMEDYYYRAEVLIGVNGPNDTSLKYVAPRKKPFPMEMPYPG
PADRRFTEVVATAGYENTPVPQGRNSRPYDGRPQCCGNNNCMPICPIGAMFNGIHSIIKA
EKAGAKVLPNAVVKFDTDENNNITALYYYDPDKNSHRVTARTFVLAGNGIETPKLLLMA
ANDRNPNNGIANSSGMVGRNMMDHPGILMSFQSAEPIWTGGGSVQMSSITNYRDGDFRREH
SAIQIGMNNTSQNHKAGVKALQMGVLGKKLDEEIRRRAACGMDIYVNHDI LANPDNRLTL
STVHKDKLGIYPHVITYDVG DYVRKAAVSSREHLMTIAKLFGATEIEMTPYFNPNNHIMG
GTIGGHDPKDSVVDKWMRTHDHQONLYIASGGVMAAAGTVNSTLSMVALSLRATDSIKRDL
QHG

SEQ ID NO.17 - Orf 3675

MNADVIVVGTGVVGCLIAEQLLDSGHSVVMLEAGPRVERWQIVENYRNLP PVSRLHFNAP
YPPEPWAPHLMSATPEQAAEYLQLEGNARAYQQGYVRYAGGATWHWAGICWRLTPEDMQ
LKTLYGVGRDWAFDYATLEPY YTRA EYALGVCGPSEPELQWPPVRSKPYPMGRLPFGPGE
QRFTDAAASIGLTNLPSAQARNSGIAYGDRPACCGNNNCIPVCPIGAKYDAATSLTRIES
KGGKIQPNNAVVKIETGADNKVQAVHYFDNNKQTHRVTGSVFVIACNGIETPKLLLMSAD
SRNPHGVANSSDQVGRNMMDQPKLVVELELAEPAWTGVGPVQGS SIMETSQGSFRSEYCG
ALFRFNNMARSRIGAMAAL EKGLVGKALDTEIRRLSACTTEIAIEHELMPDANNRLTLA
KKDWLGLPKPNIYYDVGDYVRQGSQRHSLPIARQLAKAMGATKVDISTEYTNSDHIMGGC
IMGTDPAVSVDVDCRAHDHENLFLPGGAAMTTGGCGNSTLTMSALALKAADAIHAQLGK
A

SEQ ID NO.18 - Orf 1220

MSETISTDIVVIGSGVVGSLTARKLALAGRKVLMLLEAGPRIQRDQIVSNFRHSARKDDFI
APYPNSEIAPFPDYKPEDNGYLDQTGPKDYKPEYLRVVGTSWHWAAQAWRLVPNDFRLK
SQYGVGRDWPI SYEDLEPYYYEAEILWGVSGPAEMAKYSPRKHPYPMEGVKMSYLEQRT
ARLAPKYEVL TNTTGRNSVPYDGRPQCCGNNNCMPICPIDAQYHGGIAAAAAE IAGVKLI
PQAVVYKLEHNSHGKITALHYDWNKQSHRVEAEIFVMAANAVETPRILMLSADDKNPNG
LCNNYDQLGRNLMDHPSNSATFYVDEPLWPGRGPMSPSSIQQLRDGAFRSESAAFRIDIS
NSSRVAGVTAGAIKEGLTGADLDSAILYRASHEL SIKNVLEQLPDPKNRTMLSTRKKDAL
GLPVPAFSYSFDEYIEKGMQHSLEVYADIARMLGATNVRYSTPGVYSNNQHITGTLAMGT
DEKTSVTDHV GKAW EYDNL VMVSTGVMPTVATANSTLTACALGLRTADAILGKI

SEQ ID NO.19 - Orf 2419

MMMKKPEFTPGGDASADIVIVGSGIVGGLIADRLVSQGYSVLILEAGLRISRAQAVENWRNMPF
ANRAGSDFQGLYPQSPLAPAPLYFPPNNYVNV TGPSAGSFQQGYLRTVGGTTWHWAASCWRHHP
SDFVMKSKYGVGRDWPI SYDEM EPWYCEAEYEIGVAGPSDPSMQSPSERSRPYPMDMVPFAHGD
TYFASVVPNPHGYNLVPPIQGRSTRPWEGRPVCCGNNNCQPICPIGAMYNGIHHIERAESKGAVV
LAESVVKIDTDDNNRVTA VHWLDNQGASHKATGKAFALACNGIETPRLLLQAANKANPTGIAN
SSDMVGRNMMDHSGFHCFLTEEPVWLGRGPAQSSCMVGPRDGAFRSEYSANKMILNNISRVVP
ATKQALAKGLVGKALDEEIRYRSIHGVDLSISLEPLPDPENRLTL SKTRKDPHGLACPDIIHYDV
GDYVRKGATAAHEQLQHIGSLFNGKEFNITTALNANNHIMGGTIMGKSAKDAVVDGNCRTFDHE
NLWLPGGGAIP SASVNSTLSMAALGLKAAHDISLRMKEFA

WO 2005/012547

PCT/US2004/024204

FIGURE 4A

SEQ ID NO.20 - Orf 766

MKKMTFKRLLLANTVVLACGLAGAVQAADAPNQDQLVKQGEYLARLGDCMACHTTSGRPD
 YSGGLAIKSDLGTIYSTNITPDKQYGIGNYTEQQFADAVRKGVRRPDGSFLYPAMPYPDYA
 KTSADADIALYSYFMHGVITASNSQPPQTDLSFPFSQRWGMRFWNMVFTSDKPFQPIGGAS
 EQVNRGAYIVESLGHCSSCHTPRGVAMEEKALDSSDSNFLSGGNLNGWDVPSLRGIARWS
 PDEIVDYLOSGRNDKAGVAGEMTSVVKNSTSHMTDADLQAIAYLKFLGGNPPLQAYDQQ
 KNQATTAKLTAAVDLTEGQTLYLNNCGACHFVNGLDAAAFPPQLDQASVVNAKDPQGLIH
 IILQGAQLPATEKSPSMLKMPGFGHRLSDDQVAKLATFVRQGSNDASAVTADQVKKVRE
 GLEQH

SEQ ID NO.21 - Orf 2908

MKTIFVKLLPLAIMSVIGVIGLKQAYADSNDSDADLIKQAYLARAGDCTACHTEAGGKPF
 AGGLAIRSPMGVIYSTNITPDKNAGIGSYTEQQFAEAVRKGVRRDGSNLYPAMPYPDYSG
 ITDKDIHALYVYFMHGVAPVSVKAPQTSLTFFPSLRWGMKFWNIAFASGNSYPPAPTQS
 DSADAQALSRGRYLVDLTLGHCSCHTPRGIGMQEKALNDSRFLSSGMLNDWTVPSLRN
 PDGWSVNDIAEYLSTGRNDFASVGGEMTGTVVQHSQMHNQADLHAIALYKSLPASTKQQ
 HNVKPDQLQNDTQKTVDLTLGKNLNSGQMLYLNNCEACHLTDGGGAKKIFPRLNGASIVL
 ADNPTGLISVMLKGAQTPSTANAPSVQFMPGFQRLNDQQIAELASFVRSGWGNNAPPVS
 AADVAKVRASLNTSQK

SEQ ID NO.22 - Orf 2035

MKKITLLYSAVLAGLLGCTVAQADDSSGQLVARGEYLATAGDCVACHTASGPAFTGGLKM
 TTPVGAIYSTNITPDKQTGIGDYTYDDFARALRQGIARDGRHLYPAMPYTEYAKVNDDDM
 HALYAYFMHGVTAHVQPNKPSDIPWPLNMRWPLAVWNKFLDNTPFKNDPAQSAEWNRGA
 YLVQGLEHCGACHTPRGIAFQEKASDEKGAFLTGTTLEGWHAPDLTGNVKSGLGRWSTG
 DLQTFLLKTGRNDQSAAFGSMSEAIGHSTQHLTDADLHAMAVYIKSLKSSDPEAQPPATTD
 STTAALIRGDLSTGAEYMDNCAACHRLDGKGYAKTFPTLAGNPVLLSDDPSSLISIVL
 TGGKMPVTQQSVTGLTMPDFGWRLSDQQVADVVSFIRSSWGNNAGKVEAKQVADIRKLMP
 VPNQADNPQVKAEPDPAKK

SEQ ID NO.23 - Orf 1840

MKAIKGIIIVLVLVIILLAYALWPTKTASLSPLPADNSPQLASLVSQQYLA TAGDCAA
 CHTQPGGKPLAGGLPIRSPIGVIYTTNITPDKQTGIGNYSLDDFERAVRHGILPNGDTLY
 PAMPYPSYAKISDDDVRLYAWFMHGVQPVVSQQNRASDIPWPLSMRLPLAVWRKMFAPDP
 ANTGFTADKYQSASLARGAYLVQGLGHCGTCHTPRAGTLQEKALDDSGQQYLAGGQVIDG
 WLAVNLRGDKADGLGNWTEQDIIDTLRTGHNVSHTVVGQPMAEVVAKSTSHMSDADLAAI
 AAYIKSLPAGQGSKASYTESSQTADMLARGENPTPGAQLYVDNCSACHQTSKGKVQHIFP
 AMADNPITILADNPVSVIHLILDGSRPATPQSPSALAMPGFGWRLSDKQVADLSNFIRNS
 WGNKATEVTEQQVKQVRADYPPKGENKDP

SEQ ID NO.24 - Orf 3651

MKKSILALVFGSLAFSAMAEDNSGQDLVKRGEYLARAGDCVACHTSEGGQPFAGGLPMAT
 PIGKIYSTNITPDKTYGIGDYTYDDFQKAVRHGVAKNGETLYPAMPYPSYAVVSDDDMHA
 LYAYFMQGVKPVSQPNHATDIPWPLSMRWPLAIWRGMFAPAVKPATAQPGEDPVLARGRY
 LVEGLGHCGACHTPRSITMQEKALNNSEGTDYLSGSSAPIDGWTAINLRGDDRDGLGRWS
 TSDIAQFLRYGRNDRTA VFGGMTDVVQHSLOYLSDDDINAIARYLKSLSPRDSHQPVFKA
 DDSVSQALWKGNDRQTGA AEYVDS CAACHKT DSGSYTRFFPALKGNPVVLAEDPTSLIHI
 VLTGDTLPGVQGAPSAITMPAFGWRLNDQQVANVNFIRSSWGNTSTAAVSADQVAKLRK
 SADVQGMGDASVEKLPKQP

FIGURE 4B

SEQ ID NO.25 - Orf 3053

MAKKTRRVISVVAAVVIAGALGYTAYEQYGIHKNYPQTVSLETGPALQDQIKRGEYIARL
 SDCTACHTAEGGQPFAGGYALQTPFGKILSSNITSDRETGIGGWTQEQFDKAVRHGVGSH
 GYLYAAMPYPAYSRLTDADLTDLWAYIRNLPAVNHKVVENQLPFPFNQRWTLAGWNMLFF
 KDAAFTPNPQASEQVNRGQYLVDGPGHCASCHTAKNMLGGDSSAYLQGGALQGWYAPDLT
 PDPHSGLGNWSNADIVSYLRSGSNRITASSGPMTEAVENSTQYMNDNDLNAIAAYLKSIP
 ASHPQVPTALTADDQOMVSGKKVFESQCSACHVSDGAGIRNMI PALAGNPQVNSADPSSL
 LNVVLNGSEGPFTHANPTAAGMPSFGWKLSDANIAEALTYIRNSWGNAAPAVTADQVSAA
 RKATGAKSWLGDSTIASQDSGK

SEQ ID NO.26 - Orf 2632

MKKTIIAIAVAGIVVVGALAALWMNGSTRADDVAGDQVQTSQPVSAEDSAAVKRGEYIAV
 AGDCVACHTAPGSKTPFSGGYGIDTPFGTIYASNITPDNQTGIGQWTERDFYRAVRHGIG
 RQGENLYPAMPYNAYVKVSDQDMHDLWMYMRTVKPVNQPPETHLPFPYNIRLAMRGWNL
 LFFKN SGFDANSSQSAEWNRGAYLVQGLEHCAACHTPKNMLGGDTSAYLQGS SLGQWHAP
 EITGNTYTGIGQWSEQQVVDY LKSGSNQVAVASGPMAEAVTNSTQH LTDADLRAIAVYLK
 SQPGSANQKPAALAATSPLMQQGANVYQANCSACHNSDGRGIPQLAAGLRDNPGIMAADS
 SSVITTILEGGRGAVTLNNPTSGAMPSFAWKLSDQQIAAVSSYIRNSWQNAAPAVTSQQV
 AAMRKQLKLTPQLPDNGEPAH

SEQ ID NO.27 - Orf 1750

MTIKKYIASVVGVA VVAGLGFTGWKCWHNAHQDHSFVAPASAGDTGSTAIARGKYLATAG
 DCVACHTAPGGKPYAGGLGLNTPFGTIYATNITPDKETGIGGWTDDQFMNAVRNGKGANG
 ENLYPAMPYNVYAQVSDQDLKDIKAYLDSVPAVHYTGPKTDLFPFPYNIRLMMMGNLFL
 NTAAFKADPAQSAQWNRGAYLVEGLGHCTSCHTPKNMLGADKMGVHLOGGELEGWLAPEI
 TGNTROGIGGWSDELVHYLKTGANDKTVAAGPMAEAVHNSLQHLNDQDLTAMATY LKSL
 PGSEDKSVALSGMDDVMARGQSIYQANCSACHQSDGAGVRDMVPALRGNNGLQAFEPTNV
 LHVLMIGAQAATASNPTSAAMPEFGWKLTDQQMADVSTYVRNSWGNKAPAVTASQAAAA
 RKLLSGSPALHNPAAN

SEQ ID NO.28 - Orf 1953

MMKKLMLTAGSLLLLTAGYAHADSGGDSWDLVSKGRYIAQLGDCTACHTEPGHPLFSGGV
 AIETPFGKLVGANITPD PETGIGKWT FEDFQNAMRKGHSRDGQLLYGAMPFTAYTKVTTD
 DNRALWSYLQTVQPVNRVNTNQLPFPFNIRTS LHVWDM LNFT EGEYKPD PKQSAEWNRG
 AYL VQGLGHCSTCHTPKNMLGGDKDSKFLQGGSLGVWFAPDITANTHSGIGQWTQQEIVE
 YLKTGANKYDIASGPMAEAVEHSTQYWKDEDLNAAAVYLKSLKNDSSQPQPLAADNGQMV
 NGKAIYADRC SACHVSQGGQVSHLFPQLANAPLVNAVDPASLIHVVLGSRAGGTAAAPT
 APAMPAFGWNMTDQNVADVLT YIRNSWGNAAPSVTASDVKNMRSTLEK

SEQ ID NO.29 - Orf 3687

MQKLRVFTPLAIMLAGFCGSVYADNSPASSDSTSLSRGEYLARAGDCVACHTAEGGKPFAGGLK
 MTPVGA IYSTNITPDKDTGIGNYSYDDFVKAVRQGVSKSGSTLYPAMPYASFTRISDQDMHDL
 YNYFMQQVKPV SQQNKASDI PWPLSMRWPLAFWRWTFDDKRFQ PVEGKSAEWQRGAYLVEGLE
 HCGACHTPRGIAFQEKALDQSDPVYLTGNTLEGWYAPDLTGTQSDGLGRWSQQDIVSFLKNGVT
 AQSSAFGSMSEVVHDSTSYLTDSDLQAIAYYLKSLPAAHQ TQAPASNNATAQALFKGDVSATGA
 QVYLDNCSACHRS DGKGYDKTFPSLAGNSAVLNSDPSSVIHI ILQGGQRAVTPDMPTGLTMPDF
 GWRLSDQQVADVATFIRQGWGNNAAVTASQVADIRKLIPKPASQAAK

FIGURE 5A

SEQ ID NO.30 - Orf 3820

MSRSVKVRPTSLALIIGLSVFSGKAVQAADTPSASTIIEQGKYL SVAADCGACHNSPTSG
 AAMAGGYAIA SPMGNI IASNITPSVTAGIGNYTEQQFARAVREGVNAQGDHLYPAMPYTS
 YSKMTDSDIHALYQYFMHGVQPVDPAPATKLPFPFSIRSSMALWNMLFASQQRFTPD SQ
 KSAQLNRGDYLVNVLEHCDACHTPRNFLMGQKNDLALSGGQVGSWYAPNITSDKTAGIGS
 WSDDQLFQYLKTGHVAGKAQAAGPMAEAIENSLQHLSDDDLHAI VAWLKQVPASGATATE
 SRFTQGAPSDSEAAMRATDHPDAGWVVFSSNCANCHQANGEQSQFYPSLFHNSATGAAQP
 DNLIATILFGVRRHADGQYVAMPAFGPAASFVDRLNDQQVADVANYVLKNYGNASLTVTA
 DQVKT VREGGPVPAIAYLSNPAVLAIGALIVLVILGLIVTAVRRRGKK

SEQ ID NO.31 - Orf 1956

MKQQHKLNAHKAAGFRRKLLSLCLGLSALS AVPVMAAEQVPVSQPSVDNSADALLKQGHY
 LAIAADCAACHTDPQTKKTFAGGYAIHSPMGVIYSTNITPSRQYGIGSYSEAQFEQAVRH
 GIRGDGSHLYPAMPYTSYSGLTDQDIHALYFFFTHGVQPV EQANRPTELSFPFNIREAMW
 GWNLLFLKQKPFRRDDPSQSPQWNRGKYL VANLEHCGECHTPRNTLMGSETGSAQYSGAAL
 GSWFAPNLTSDQQSGLGSWQRDQLITYLKTGHVAGKAQAAGPMAEAVTNSLQYLSDDDIG
 AIVTYLQSLPPVSEPDQAKATGDFGSSAGNSSDSEVRGTQPMGSVLPDDITGKALYDTTC
 ASCHQSSGAGTTDNFYPSLFHNTATGGNTPNNLVSAILFGVQREVNGKQVLMFAPGPGSD
 VQSLNDEQVAKLSNYIFKQFGNPQLSVTADQVKT LREGGPQPF LAKYAASGSAVGGVILL
 LIIVLIIVRISRKRR

SEQ ID NO.32 - Orf 2446

MKCAYLSLLISTLLYAGFSPATQAETPATAETLLAQGKYL SVAADCSACHDSPDHHVMAG
 GNSINSPLGKIVASNITPSVHYGIGSYTEQQFSDAVRKGINAQGENLYPAMPYTSYSQLT
 DSDIHALYFFFHMGVTAVDRAAGATQLPFPFNLRISMKLWNALYADNKPFRPSSSQTDQV
 NRGNYLIYGLAHCDCHTPRNALMAEKSDQSLSGGSLGQWYAPNITSDKSSGIGNWSDQQ
 LYQYLKTGHAVGKAQAAGPMAEAIHSLQYLSDDDHLHAIVASLRLTRPVNTASADRGMQG
 KAISDENSIRGTKVASGEPVSGPMSGAILYSGNCAACHTPSGAGSYSQNYPSLVHNTTVG
 STDPTNLIATLLFGVHRTVDQQSITMPAFGPQGYTDRLSFAEIATLATYVRQTYGAGGEA
 VSEQQVEQVYQGGPKPLIGWLADGRIQALIVVVL LLAGLIITVVRKGRKA

SEQ ID NO.33 - Orf 3674

MKKHAIKFSLSLMFAGSMLWAGSAAAATGDAAA AISRGEYLATASDCAACHTDKGGLPFA
 GGLKIESPVGTIIASNITPSLTAGIGHYTEQQFADAVRKGIRADGANLYPAMPYTAYSVM
 TDQDIHDLYQYFMQGVKPVDPHPAAETELPFPNMIRMMKAWNLLFLNDKPFSPDASQSAA
 WNRGKYLVTGAHCSTCHTPRGPLMEEESSQFLSGGQVGAWYAPNITSDPQSGIGRWSQA
 DIVQYLRTGNLPGKAQAAGSMGEAVEHSFQHLTDDDLNAIATYIRTVKPVATPENAGSRF
 MQGDSDATGKIRGLSQQQVTDKQOGLALFQGNCASCH EAGGQSRDSYYP SLFHNSVT
 GAENSNLIATILNGVNRTRTDGQVFMPGF GHHPNDINNLTDEQIASLAN YVLTTYGKPS
 KPVTAAAMVATVRQGGPGSSLVLLARFGIAAGVVVVLILLGFWVVRKKNVRDPS

SEQ ID NO.34 - Orf 1221

MKKLLSLCIAGALAGIMLNSAAMAEDSNAQSLIAKGQYLSVAGDCAACHTTSGGKPFAGG
 LAIATPIGKIFSTNITPSKTSIGIGDYSLQEF EKAVRQGVKDGANLYPAMPYTSYAKISD
 EDMQALYAYFMHGVAPVDEKGPQTALPFPFNIRLSMAGWNLI FAGDKPFTPD SNQSAEWN
 RGAYLVQGLAHCSTCHTPRNALMAEESGQALAGASLGTFAPNITPDAHAGIGKWSASDL
 ATYLSTGRSPNGSQAGGPMLAIDKSFSKLSQSDINAI VTYVRSVKPQSANAAPGQVPAS
 APVVSDFALMNGTASDGAKLYEAHCSTCHQASGQSGNGLPALYGNAAALHRPVADNAVMAI
 LDGLTPTQGQAMP SFKTAMNDQQIATLTNYLFKTFGDAGVQTTADRVKVLREGGAPSPLL
 AIAKGGMIAAVIVVLLLIIVGGVMVKSRRKRR

FIGURE 5B

SEQ ID NO.35 - Orf 1984

MKKYSALLTLSAAFLFSPLALAATSSNSDLVSRGEYLARAGDCTACHTAAGGAHEYAGGYK
 FNMPMGITIVAPNITSSVQYGIGNWSEADFAKAVRQGVPRPDGSHLYPAMPYTSYATVTD
 MQALYAFFKTVPAVDKAPADKNDLKFPFNLPGLMGIWNALFASDAPFKADPALTAEQNRG
 KYLAEGLAHCSTCHSPRNQMMEDTHQLLAGNHVDGWLAPNITSDAVSGIGGWSQQELTE
 YLKTGHVEGKAQAGGPMADAIHSHLSDSLASIATWLKTVPAIRTPGQTQPSWAAAP
 ASKVDWTSYQTGGGKNNSPAYRDSSTTDGAVLFDSSCAACHQSSGQGSDDHYFPSLTHNS
 AVGAADPSNLVMAIVDGIHRKTPEGEAVMPAFSSETQAIHSWLNNDQIAAVTNYVTEKFG
 HGNAGLTGADVEKIRNGNSNPFLIKNAGGLTIGGIVIVVIIIIALLAARSRKRR

SEQ ID No.36 - Orf 3400

MKAVIIRSAIALALMHGSLALAADDNADLIKRGYELATASDCTACHTAPGGPAYGGGYPV
 ATPFGKIWGSNISSDKQFGIGSWTDDQFVAVRQGVGKNGEQLYPAMPYDAFTKMKRDDV
 LAIKAYLMSLPVHKAAPETSLPFPFNQRWGMRFWKMFNLTEGELKNDPQQSPQWNNGRY
 LVEALAHCTTCHTPRNLTMGMDTSKPLSGGDLGDWIAFNITPGKSGIGDWSSQDIVTYLK
 TGYLAGKASASGPMAEAIHSLQYLPDSDLQDIATYLSVKPVDDEKQSVPRDSQGQPSD
 AIIRLRGADAATLQSQPGAVVFEGNCSTCHGAEGAGSGQGFAHPSLFFHSSTGAIDPKN
 VVSVILNGVNRHMQQGDI FMPSPAPQLNDQQVADVANFVMQKFGNPAAEKVDTSQVSKAR
 KNASLPLPPTFADGANP

SEQ ID NO.37 - Orf 3041

MIRSSFKRSRNFLPLAGLLFCAAGYAQTGSAQPDVATQPTPTQPAAGTQGTTLIQQG
 EYLAKAADCEVCHTATGGQTFAGGLGFKTPFGTIFSSNITPDKTHGIGQWSEKQFSDALR
 YGIRADGKNLYPAMPYTSYSKLTADIAHMYAFFMSLKPVATDPPENKMGFPYNQRIALK
 GWNLINFHYQPFKQDPDQSAEWNRGHYLATALGHCEECHTPRNLMGLSDKSYAGAMVDG
 WEA FNISDNTSGIGRWSHADLMQYLKTGSGVPGVATTGGGMADVISHSLRFLSNDDLSAL
 ATYIKSVPPQKTASQNRSGYGDNVQSDITQAVRGMPIDDSAPSGAVLFNGNCASCHGKTG
 QGIGENRYYPSSLNNSVVGADKANNLVQVILYIGIDRTNGKGEHIVMPGFGDELTD SQIAT
 LTNYLRTNFGTNPAPVDAAQVKALRENNVMVIPGYLLILGGVIGVILVAIIMYFRRRKA
 ARNHAG

SEQ ID No.38 - Orf 2420

MKRFSRVKLTLGLLGGGLTSLAANAADIDQALLQOGEQVATASDCQACHTAPGSKTAF
 SGGYAIASPMGAIYSTNITPD PATGIGKYTEQQFIEAVRHGVRADGAQLYPAMPYTSYR
 MMTDSDIHALYFFMHGVKPVDDQNTETQLSFPMRFSMKFWNLLYADTKTFQQDPQK
 SAEWNRGNLYLVNGLAHCDTCHTPRGFMNEQTDQPLAGAPLGSWYAPNITSDKVSGIGG
 WSNDEIVQYLKTGRAAGKNQAAGGMAEAVEHSLQYLPDSDLQAIATYLKQTTPIRTPGE
 TQAAYSYGSSSTNVDDQVRGMAPNNARDSLTSGAALFSGSCASCHQPDGAGSKNQ TYP
 LFNNTATGMIHPQNLIATILFGVQRNTKDHQVLMPGFGASTSYVDSLTDQQIADISNYV
 LHN YGNPAVTVKAGDVAWVRKGGHPPALVALQPYMIPAIAGV VIIIIILLVAFRLRRSR
 RKS

FIGURE 6A

SEQ ID NO.39 - Orf 3653

MSEQNKGQSRDFFLLKTITLAPAMAVGSTAIGSLALSPAVQAADTQTSGPQKARDYQPNW
 FTKEEFAFITAATAVAKLIPADSRGPGALEAGVPEYIDRQMDTPYATGSNWYMQGPFAPDTP
 KELGYQLPLVPRQIYRLGLADADNFCKQQYGHVFAELSDDQQVTALKAFESGQAKFTQLP
 ATLFFSYLLQNTREGFFSDPIHGGNQGMAGWKLI GFPGARADFM DWVERGEHYPPFPVSI
 RGERA

SEQ ID NO.40 - Orf 3689

MKNTPRSKDSTGRRLFLQRSLSLIPLVAATGTPFATSQAAEKKTPAVTQDYVPQFFDPQQ
 WAFINAAVDRLIPEDQNGAGAVSEGVPVYIDRQMELPYGYGHLWYMQPPFASHSDPTLGY
 QSPLVPRELYRQGI ALTEHYCQQT FHK SFAQLTTDQQDQVLQLLEKNTLTDNNLSGSLFF
 EQLLDNTKEGYLADPVHGGNQT LASWKLIGYPGARADYTD TVAQPNVPYPLGPVSI SGKR
 SV

SEQ ID NO.41 - Orf 1842

MSDKPSHSRRDFFLLKSLTLIPAVSVGGAITSGIAGPGNAQAAETSATAATAQTPYSPVFF
 KPDEWAFVKAACARLIPADDMGSGALEAGVPEFLDRHLQTPYANGSVWYTQGPVEAGPE
 FGYQGRKTLSEIIRSGIRGVIGWTQSNKQOTFDALTHAEQEEILVALEKGKIHLEEMDAK
 TFFDYFLGEVRNGFFADPSYGGNKG MVGWKLIGFPGMRADYIDFITVRDKPYPLGPVDLA
 GNRG

SEQ ID NO.42 - Orf 2037

MKENSQPPAASRRKFLQTALAIIPSTALATSVVPAALAAEQTKNPTRDYVPVFFKDDEWR
 FIIAATDVLI PGDEYGP GAVSEGVPV FIDRQMEMPYGYGQLWYMKPPFQEGSPLLGYQKN
 LTPRDIYRRGIAALNKACQTTYQHPFASLATADKVQVMEDLESGKLVTEVDVGKLFFAQL
 LENTKEGYLADPIHGGNQT MASWKMIGFPGARADYVQVMDNPGKPYLPGPVSI SGKYGA

SEQ ID NO.43 - Orf 2910

MKQSGIGRRPFIIGSLIGIASLGMKCGVSSVFAAVTSPLDELNSYQPVFFKPEEWQFIMA
 ACDRLIPQDEEGPGALET HVPVFIDKQMLTPYKGEDWYMEGPFNAHASTLFGYQLPFPL
 QVMYQRGIKATNSYTRLHFNQDFAALTAQQDAVL SALEENKITFSEFSEPDLSASYFFT
 RLLENTKEGYLSDPKYGGNKGMAAWVMINFPGARASFPTWIKIHNVKYPLGPVALNGDVA
 QSS

SEQ ID NO.44 - Orf 1951

MSDPSSKGISRRRLLSGSAAGLTVAAVSSANATTITGIPRWMLFDHNSPITPTSPGLKFL
 TQEEATEVDIAIVSQLIPADELSVSGKDAGCTVFIDRQLAGSYGDASRNYMRGPFREGTPA
 QGDQSPLVPRERYRLGLAGLSDYCQQKYQKLFSQLDSATRDEVLTGLEQGKINLTGISGK
 MFFDQVLTNTMEGFFSDPVYGGNRNMVSWKMIGFPGARYDYRDYLTQKLDLVPISIM
 GSTAWNKV

SEQ ID NO.45 - Orf 2634

MKRREFLSSMAAFGAASAIPLTNAAEISGGQPWPPGQVSLPPGLPRKGGLQFFTRHQLET
 VGAI AERFI PADELSISGKEAGCAIFIDRQLAGDFGQAVTVYRLGRFVKGTPEQGPQSPL
 TPADQYRLGLNALDSYCQQQFHNFTELTDGQQDQVLQGMETGKISLAENFDSKVFFELL
 LQNVREGFLSDPLYGGNKDMASWKMIGFPGARYDFRDVIAKKGQKLNIIPTSLIDNNL

SEQ ID NO.46 - Orf 764

MLLQKNTTRRKFLLGSLMALPLTELVLKGLTAAQAADMAAPELTSYKPAFFTADEWQFIL
 AATDRIIPAGGPGKAPGALETNVPFI FIDQQLHDEHFGKEIYMEGPFNPHAPATMGYQVPL
 YPQQIYQTGIRLTNQWSQQNLQKPFHQLSEADKDKVLTGLQKNTLDF AALGENTLKGS LF
 FSQLLGETKHGYLADPMYGGNKG MKAWIAIGFPGARASYLEWVKQHNVKYPLGPVSL LGE
 TA

FIGURE 6B

SEQ ID NO.47 - Orf 3051

MQRRKFIKTGLILAGTGTAASVFKPAGAAARDNILNGGKLWKAKETPPPTPADPTKRLYL
TEQEYAIQITAI FNRLIPADELTVSASDAGCVVFIDNQLAGNYGKASWRYNVGPFENGTPS
QGNQQPYTPAQIYRIGLAEIEKDCQSKFSKSFSELTNDQQDKYLEQMEADQIKYPTLSSK
DVFSQFLSNVQEGFLADPIYGGNRNMIGWKMIGFPGARYDYRDYAPLKGTKLNIEPVSII
QLLKA

SEQ ID NO.48 - Orf 1748

MKRRRFLASLGVLLISTALKVKAKIISGGMPWVVHAVKPPQPVVAGEWQFFTPEEVAIIE
AIADRIIPQDELSIGGKEAGCALFLDRQLAGDYGKAVSIYRLGPFIONGLPEAGPOYKDV
PAERYRLGLASVNEISQAKYNGKKFNEISEEQDDLLGKIESGVLPLTGVDGKLFFDQLV
INMREGFFADPLYGGNKDMAGWKMLGFPQAQYDFRDVIDKRGEELNIKPVSMVTNNDQS

FIGURE 7

SEQ ID NO.49 - Orf 3819

MTANNRHPSGVSRRLQLQGMGILSVAGLCGSLFPSFRAAAAELQDSGFIPLSEFLVNRRV
 NPILAQRYYDALHRHDEKFDQKLALLKQDIQPGKYQNIDDFLQKNAVGTDLRQAAGQVIS
 AWYTGTVGNDEKLELIAADAMMYVPTSGVLVVPTYGSGPISWAAVDNKPAPHQGPAV

SEQ ID NO.50 - Orf 2447

MKLTDTISTDRRKLIKSLSLTVFSVSGRLVTCPAFAGGLPASADFHEFSTFVIGRPVDPVI
 SGRYFAALQAADGHFIQQLNQAMVASVPFRSQGIDTMLASLPHDSDFNTLKKITSAWYLGIV
 GEGAGATLIAFHDALMFQPTREYVFPVPGYGGGPDSWVSLKHPDLLSEDTEQEQKNG

SEQ ID NO.51 - Orf 3676

MKNEIIRDDSPAENLSRRKVLLGGLILLGSSYLGPSPAWADTLNDQATIDQFMQLSQL
 LVNHQLDPVTGQRLAAAMISGNMITRQQITSLLAVAQARQAKVVEDFFSDIPQGELKNAA
 LSIISAWYKGVLDAPGAEVFAYEKALMYQPTIDVMTIPTYAITGPNGWSSHAAPLADMPDF

SEQ ID NO.52 - Orf 1219

MIDMLNMISRRRILOGMGALAATLLPSGILPAFADTPANSDFNDISRLLTGRNTLSAEF
 SSALFSAFTKIDSFRPQQLARLKQWITANSVPAADLQKRLTADSSVADLAGLPALILTGW
 YLGIAGSGDKAVCVTYVDALANQEVASVLNPPTYAYGAYGSWATKPF

SEQ ID NO.53 - Orf 1957

MNNHNAPETQPELSEGLRRRKLFGQTGGLVASFAIGSAIAGSTLSNGANAATTSAGPDT
 QTLNQFMKTSRLLTGHQNLDTLGLQRLYVAFSEKDPQFITQLSALNQWIADKQPADVEAL
 DSQLSGQPLHALMMSVIKGYLGVIDDSHHAKVYAYQNALMYQVPRDGMVIPTYAHNGPD
 YWTADPPPVDRLNF

SEQ ID NO.54 - Orf 3043

MNKATPVSPGERRRFIKLLAASTVAGTVSSLLPGQIAWAIDAGQPAVAGFPFMTVSEII
 CGYPTLDNALGKRIFSLISAEGDASQSIAELQKQLNADMSSAEMQAALKTLDTPAQQLF
 SEILRGWQIGIVGSGKQSQVVAYEYALMYAPISDVVVLPTFARGEPHYWAYPPVIKTGKL

SEQ ID NO.55 - Orf 1982

MKFVIDQESDTGEISASRRSFLIKITALLASFTLIPAHAVITTPADV GASVISQLQTTAQ
 FLTESQQDPQLIIRAANALLKVNSNFAGDLQQLSSLIADNHIANLKD LKTSNLFEGKPQQ
 TAKDILSALYLG YAGTPVMLSSEDNVV FVAYA QARTYQLTKDFTPVPSYSRWKSGYWAHL PAG

SEQ ID NO.56 - Orf 3398

MNLTRRRLTGSAGLIVAGVLSQTLSGRYALASPPLASAVAPSAGFNTLSVLITGQDKPD
 ALLAQRLYSWLAAHTSGLDSQLETLSLLQQHSDANGSTLLSLMKSQPENINTLYQSLVS
 GWYLG VVGPLPRPDCIAFENIVSYQLLKQSVLPPSYAPGQPGFWVQPPAGRVHV

SEQ ID NO.57 - Orf 2418

MKQIFEQSHTDLPENGTGSSRRGFIKSALVLTASGLVASLPLRSFASSVVHGGDTTQDFI
 SVSQAITEHKHINPQLAAHFLSAFIKRDNQFSSKITRLAQLYQTGDTAIVFKNKAVAAGL
 GDFLQQILTAWYTG TIGDDYKGT LVAYKEALMYDTVSDGLVVPTYCGNGPLWWTVPVPDP
 LDPELINNL

WO 2005/012547

PCT/US2004/024204

FIGURE 8

SEQ ID NO.58 - Orf 3397

MKTKTLLAAALLTTGYAATAGAATLTNSQANHVIENAESTIKAQNSTGCAVVVDNDGMLLSFQ
RLDGATPGCIDAAIGKARTSALYHAPSVKFMQRLQSGETTVLAIPHAVPLGGGYPLTLQGEVVG
AVGVSTPKQDLDNQASETAAKSLK

FIGURE 9A

SEQ ID NO.59 - Orf 765

ATGGCTCAGATAACTAAGAAAGAAGTCGACGTCGTGGTTTGCGGATTCGGCTGGGCCGGT
TCACTGATGAGTATTGAACTGGCCATGGCAGGACTGACAGTGCGTGCACTGGAAAAAGGT
CCTGAACGTGATTATGAAGCGTTTGCCATCCTAAACCGGCAGACGAATATGCCTATGCG
GTTTCGTAACAAGGTGATGACCACCCCGCAGATTCTGCGGTGACCGTGCGTTACACCATG
CAGGACACCGCACTGCCAACCCGTAAGTGGGGCGCCTTTGTGCCTGGCGGCGGTGTGGGT
GGTGCCGGTATGCACTGGACCGGTGTGTTACTGCGTCCGACCCCGACCGATATTAACTC
AAAACCTATGCCGATGAGGCCTACAAACCGGGTGTATTACAGGAAGATATGCGGGTGCGT
GATTTCCCGTTTACCTGGAATGAGATTGAGCCATGGTTTGAAAAATTTGAGCATATCTGC
GGATTATCCGGTAACACCGGTAACCTGCGCGGCCAGATAATGGAAGGCGGAGATCCGTTT
GAAGGGCCCCGTGCCAATCCGATGCCATTGCCGGCACTGGAAAATACGCTCAATAACGTG
ATGTTTGGCGATAACCGTGAAAAAATGGGTTATCACCCGTTACCATTCGTCGGCTGCC
GCTTCCCGTGTCTGGACCAACCCTTACGGAAACACTATTGCTCCGTGTAACCTATTGCGGT
TACTGCTCTAAATATCCGTGCCTCAACTACTCAAAGCCTCACCCAGACCGCAGTGCTG
GATTCAGTGAAGCAGATGGATAACTTCTCTTATGAAGTAAACGCGGAAGTGTTGAGAGTG
GTGCTGCATGATGATAAGAAAACCGCCAAAGGGGTATCTATATCGATGAGCAGGGTAAC
GAATGCTTTCAGCCGGCAAAAATTGTCATCCTCAGCAGCTTCCAGTTCTATAACGTGCGT
CTGATGCTGCTGTCCGGTATTGGCCAGCCTTACAACCCGGTCACCGAAGAAGGGGTAGTC
GGACGTAACCTATGCGTTCCTGAGTAATGGTTCTGCGACTCTGTTTTTCAAAGATAAAAAC
TTTAACCCGTTTGTGAGCTCCGGACCGACCGGTATGCAGTTCAACGATATCTCTCCGGGC
AACTTCGACGGTCCGGGACTTGGCATTATCGGCGGAGCAAAAATTCAGAGTGCCAGTCG
ACCGGAACCCCAATCAGCACCGCGCTGCCGCCGGGTACTCCCTCCTGGGGAGCGGGCTGG
AAAGAGGGGCTGGAAAACCTGGTACGGCCATTCAATGAAGGTGGGGATCACCACTTCCTGT
ATGTCGTACCGTGACGTTTACCTGGATCTGGACCCGACCTATAAAGATCGCCACGGTCAG
CCATTATTGCGCATGACCTTTAACTGGAAGCACAACGAACTGCAGTTACAGCAGTACCTG
AAAGGCATTGTGCGCAATATCGTCAAAGAGATGAACCCTGACAGTTTCAGCATGAGTTTC
CTGCCGATGGGCGCTGACTTTGATCTGACCAAGTATGTCTCTACCCATAACGTGGGCGGG
GCTATTATGGGCGATGATCCGAAAACATCGGCCCTGAACCGTTACCTGCAAAGCTGGGAT
GTTCATAACGTATTTGTGCCGGGAGGTAAATGCGTTCCCGCAGAACTTCCAGTCCAACCCG
ACCAACACTATCGGTGCAATTACGCTGATGGCAGCTAATGCCATTAAAGAACAGTATCTG
AAAAATCCGGGCCCCAATGGTACAGGTG

WO 2005/012547

PCT/US2004/024204

FIGURE 9B

SEQ ID NO.60 - Orf 2909

ATGAAAATTACCAATGAACCTGTTGATGTGGTCATCGTCGGTCTGGGTTGGACCGGTGCT
ATTCAGGGTATTGAGTTAGCCAGGACAGGATTAAAGATACGAGCACTTGAGCGTGGGGCC
GATCGTACCAGCGCCGAATTTGCATACCCGGTTCCGGCAGATGAACTGGCTTACACCAAG
CGCCACAAAATAATGCAAAGCCCGGCGGTTGCCGCATTCACCACACGACATAATCTGAAT
GAAGTTGCGTTGCCAATGCGAGAACTGGGTTCTTTTAGGCTTGAGATGGTGTCCGAGGC
GCAGATTTACACTGGACAGCCATGATTACCCGGCCAACTCCGGTTGACCTGAAACTGGCA
ACCTATGCCCCGAGAAAAATTTGAAAAATCTCAGCTGGATAAAGAACTGAGGATTTATGAC
TTTCCGGTAAGTTGGTCTGAGATTGAACCTCATATGGACTTTTTTTGATCAGGTTTGCGGA
TCATCAGGTCAGGCAGGCAATGTCCGCGGCCAAATCCTGCCAGGGGGTGACCCTTTTGAA
GGTCCTCGCTCCAGTCCATTTCCAAATCCACCCTGATCGATACTGAATAGCAGTATG
TTTCGCCAGGCGGCCACTGAAATGGGATATCATCCCTATTCAATCCCTTCCGCTGCCGTT
TCTCAGGCATTCATAACCCCTATGGCCAGCAGATTGCTCCTTGTAATACTACTGTGGTTAC
TGTCAGTTTTATTCTGCCTTAATTACTCAAAGGCCTCACCTCAGACGGCAATTCTGGAT
CGCCTGAAGCAGTATGATAATTTTGACTACAAAACCTCATGCCAACGTTATCCGTGTAGAA
AAACATGCAGATGGCAAAACTGCAACAGGGGGTAACCTATATCGATGAAAACGATAATGAA
GTTTTTCAGCCAGCAAAAATCGTCATTCTGGCCAGTTTTGGGCTGAACAACGTACGTTTG
CTGCTAAATTCTAAAATTGGTCAGCCGTACAATCCAGTGACCGAAGAAGGGGTGGTTGGA
CGTAACTATAACCCACCAGTATGGTGGTGGTATCACGCTTTACTTTAATCAACTTGAATTT
AATCCATTTGCAACTGCAGGACCGACCGGAGTCGTTATTACCGATTTTGGTACCGGAAAC
ATCAATACTGCAGACCTTGGTTTTATCGGTGGTGCCAAAATCTATAGTTCGCAGCCGACG
GGAACCCCGATGGGCGCGCCGGTGATTGATTCCGCCGCTAAGTGGGGAAGTCGCTGGAAG
AAAGGCCTGAAACAAAGCTACGGACATTCAATGGCCATTAAGCTGGAAGGCTCCAATATG
GCCACTCAGACCAATTATCTTGATCTGGATCCTAACTATAAAGATAAATTCGGTATGCCT
TTGCTACGCGTCACTTATGACTATGTGCAAAATGATTTACGCATGCTGCAATTTATGCGC
GAGAAAATGGTTGGTATCGCTGAACATCTAAAACCAGACCATTATTCCGTTGGAATGCTA
AAAATGGATAGCCATTTTGCCAGTTCTCCGGCTTATGCTAATAACCCATAATGCAGGTGGT
GCAATCATGGGAGATAACCCGAAAACCTCAGTCGTAAATCGCTATTTGCAAAGCTGGGAT
GTGCATAACGTATTTGTCATGGGTGCCTGTGTGTTTCCACAAAATGTCTATGCTAATCCG
ACAGCATTGGTTGCAGGGCTGACTTACTGGTCAGCTAAGGCCATTTCGTGAAACGTATCTA
AATAACCCCGGTCCGCTGGTTCAGGCA

WO 2005/012547

PCT/US2004/024204

FIGURE 9C

SEQ ID NO.61 - Orf 3688

ATGGCAACAACATAAAAAACCGGCAGCCGACATCGTCATTGTTGGCTTTGGCTGGACAGGT
TCTCTGATGGCCCGTGAATTAGCAGATTCCGGATTAAAGATCGTCGCGCTGGAACGAGGT
GAACAACGAGATACTTATCCGGATTTTGCTTATCCACGTATAACCGATGAACTCACCTAT
GGCATTCTGTCTGAAATTATTTTCAGAACGCAGCACGGGAAACCGTCACCGTCCGTCATACT
TCGTCCCAGACAGCACTGCCTTATCGGCGTTTTTGGGTCCTTTCTGCCAGGTAATGGTGTG
GGTGGCGCAGGTGTCCACTGGAATGGCATGTTGTGGCGTCCACTGGCCGCCGATTTAAAA
ATGCACTCAACTCTGGTTGAGAAATATGGCGCGAATTTTATTCCGCAGGATATGACCGTC
CAGGACTATCCTTTCACTTATGAAGAGATGGAACCTTTTTTCGACAAATTCGAAAAATA
TGTGGTGCTTCCGGACAAGCTGGCAATCTGAATGGTGAAATTCAGTCAGGCGGAAACCCG
TTTGAACAGCCCAGACAAAACCCCTACCCGACCAAACCACTGCAACGCTTGTATGCCGGT
GATGTGTTTCGCCAAAGCGGCAGAAAAAATGGGTTATCACCCATTCCCTTGCCCTGCTGCT
AACTGTACCGAAGCCTGGACCAACCCTTATAAAGTGCAACTGGGAGTATGTAACCTATTGT
GGTTTCTGTGAACGCTTTGGTTGTTTCAATTATTCCAAAGGCTCTCCACAAAGTTGCGTA
ATTCCGTCACCTCAAGGCCTATGATAATTTCGAACTGCGAACGAACGCCAGGTGATTCCG
GTCAATACCGATAATACGGGTAAACAAGCTACCGGAGTGACCTATATTGATGGCAGCGGT
AATGAAGTGGAACAACCTGCATCCCTGGTGATTCTTAGTGCTTTCCAGCTGCATAATGTC
CGGTTGTTATTACTCTCTAAAATCGGTAAGCCCTATGATCCGCAAACCGGCGAAGGTGTT
GTCGGGCGAAATTATGCTTACCAGATGACAGGAGGTTCAAACTGTTTTTTGGACCTGAT
CAGGATTTTAAACCGTTCGCGGCTACAGGTACCACAGCAACCTTTATCGACAACTTTAAT
GCCGAAAACCTTCGACCATTTCATCGCTTGGATTGTGCGGCGGCTCAACAATTTCTGCAGCA
TTCAGTGGTGGCCGTCCTATTCAACAGACACTGCTACCCTCTGACGCTCCCCGTTGGGGA
AGTGGCTGGAAAACCGCAATAAAAACCCACTATGCACATACCATGTCTATTGGTGCTTCA
GGTTCGGTAATGCCCTACCGGCAATGTTATCTCGATCTGGACCCGACTTATCATGATGTT
AACGGACAACCATTACTGAGAATGACGTTCTGACTGGCAACCTAATGAACTGAAAATGACT
GAGTTTATTGGCGGAAAAGTTGAAGAGATCATCAAGGTCATTAATCCGCCACACTATGAA
ATGGGCTTTATGAACATGAACAGTCACTATGATGTTTCGCCCTTATCAGTCAACACATACT
ACCGGCGGTGCAGTCATGGGCGACTCCCCCGCACCAGCGTCGTCAACAAATACCTGCAA
AGCTGGGATGTTCTAACCTGTTTGTCTTGGAGCCTGCTGTTTCCCGCAAACCTGGCC
TACAACCCGACAGGTATTGTCTGTGCTACAGCATTGTTCTCGGCACATGCGATTAAAACC
CGCTATCTGGCTGCACCTGGCCCGCTGGTTACAATA

WO 2005/012547

PC 1/US2004/024204

FIGURE 9D

SEQ ID NO.62 - Orf 2036

ATGACAATTAAAAAAGATCCGGTAGATGTAGTTATTGTTGGTTTTGGCTGGACAGGCTCT
GTGATGGCAATGGAACCTGGCCGAAACCGGCCTGAAAATTGTTGCACTGGAACGTGGTGAA
CAGCGGGATACCTATCCTGATTTTGCCTATCCACGAATTGTGGATGAGCTGACTTACGGT
GTCCGGTTAAAACTATTCCAGAACCTGTCTAACGAAACGGTGACGGTGCGTCATGCACCG
GGCGATCTGGCTCTGCCTTATCGGAAAATGGGGTCTTTCCTGCCAGGTGACGGGGTAGGC
GGCGCAGGTGTTCACTGGAATGGCCTGCTGTGGCGCCCTCTGGAAACCGATCTGCGACTG
AAATCAACGATCACCGAAAAGTACGGAGCCGCGTTTATTCCACAGGACATGACGTTACAG
GATTATCCGTTTACCTATGCTGAAATGGAGCCGTTCTTTACCCGTTTTTGAAAAAATCTGT
GGTGCTTCCGGCCAGGCCGGCAATATTAACGGTGAAATTCAGCAGGGGGTAACCCGTTT
GAAGCCCCGCGCAGTGGTGCCATCCGACCAGCGCCCTGAAAAGCCAGTATTCCGGAGAA
CTGTTTGGCAAAGTCGCCAAAGAGCATGGTTATAGTCCGTTCCCGGGGCCTGCGGCAATC
TGTACCGAATCGTATCAGAATCCGTACGGGGTCCAGTTAGGGGTGTGTAATACTACTGCGGT
TTCTGTGAGCGTTTTTGGTTGTTTCAACTACTCGAAAGCTTCTCCGCAAACCTGTGTTATC
CCTGCGTTGCGCCAGCATACTAATTTTGAATTGCGCACTCATTCTCATGTAATTCGTGTT
AACAAAGACAGCACAGGGAAAAAGCGACCGGAGTTACCTATATTGATGCCAATGGCCAG
GAAGTGGAGCAACCGGCGGCCCTGGTGGTCCTCGGAGCGTTCCAGCTGCACAATGTTTCGT
CTGCTGTTGCTGTCAGGGATAGGTCAGCCGTATGACCCTCGTACCGGCGAAGGTGTGGTC
GGCCGAAATTATGCCTACCAGGTCAATGGTGGTGTGAAATTGTTTTATGACAAAGACCAG
TATTTTAATAATTTTGCCGCGACCGGGTGCTCAGGAACCTATATCGATAACTTTAACGGT
GAGAATTTGACCACTCCTCACTCGGGTTTATTGGTGGAGGCACTATTTCCGCACATGCC
ACCGGCGGCCGTCCTATTACAGCAGACAGAACTGCCTTCCGGCAGCCCTAAATGGGGTACC
GGCTGGAAAAAAGCGATGAAAGATAATTATCTGCACAGTATGAGTGTTGGTTCAGCTTCC
TCAGTGATGCCTTATAAGCAGTGTTATCTGGATCTGGATCCTACCTATAACCGACGGCTAC
GGTTTGCCTTTGCTGAGAATGACGTTTGACTGGCAGGAGAATGATCTGCGTGTGACACAG
TTTGTTGCCGGTAAAACCGAGGAACTGGTCAAAGCGCTAAAACCTCGCAGTTATGATATG
GGATTCAAGAAGCTCAATACTCACTATGATGTCCGTCCTTATCAGTCCACGCATACCACC
GGCGGTGCGGTGATGGGGGATAACCCCTCGTACCAGTGTAGTGAATAAATATCTGCAAAGC
TGGGATGTGCCAAATGTGTTTGTCTTGGTGCCTGCTGCTTCCCGCAGAATATTGCCTAC
AACCCACCGGCATCGTAGGGGCAACCACACTGTTTGCCGCGCATGCGATTAAAACTCAG
TATCTCCGTAACCCGGGACCACTGGTACAGGCC

SEQ ID NO.63 - Orf 1841

ATGGCTATTGTAAAAAATAAGACCGATGTAGTGATCGTGGGGATGGGTGGACCGGTGCC
ATTATGGCGAAAGAGATGACAGATGCCGGGCTTTCGGTAGTCGCTCTCGAACGCGGAGCT

FIGURE 9E

GATCGGGATACGGAACCAGATTTTGCTTATCCTGGAGTCGTAGACGAATTACAGGGTTCG
GTGCACCGCCGCTATCTGCAAAGCCTGCACCAGGAAACGGTTACTGTGCGTCATAACACG
GGTTCTGTGGCGGTTTCCTTACCGTCAGATGGGCTCATTTAAGCCTGGCACCGGAGTGGGA
GGCGCCGGCAGCCACTGGTCAGGATGCCATTTTCGTCCGCTGCCTGAAGATTTACGCCTG
CGTAGTAACCTGGAGGAACGTTACGGCAAATCTTTTATTCCGTCAGATATGACTATCGAC
GATTTTCCGGTCAGCTACGATGAACTGGAACCGCATCTCGATATGTTTGAAAAGGTTTGT
GGTACCTCAGGCAAGGCCGGAGTGATCCGTGGAGTTGTTTCAGGCGGGAGGAAATCCGTTC
GAAGGCTCGCGTAGCGGAGAGTATCCACTGGGCCCTAACCCCAACTATCTGGGTGCCGAG
TGGTTTTATAAAGCCGCCAGGGAAAAGGGCTATCACCCCTATCCTATCCCGGCCTCCAAT
GCCGCCGGCCCTTACATTAACCCTTATGGCTGTCAGATGGGCCCTGTAAATGCCTGTGGT
TTTTGCAGTGATTACGGTTGTCTTAATTATTCAAAAGCGAGTCCGAACATCTGTATTATG
CCGGTTTTGCGTCAGCGTAAAAATTTGAATTACGCACCCATGCTCAGGTTCTGAAAGTA
AATCTGAGCAGCGACGGCAAAAAGCTACCGGAGTAACCTATCTGGACAGTAATGGTCAG
GAAACCGAACAACCGGCTGACCTGGTGCTGTTATGCGCCTTCTCACTCTATAACGTCCAT
CTGATGCTGCTGTCACAAATTGGCAAGCCTTACGACCCTGTCAGTAATGAAGGAACCGTA
GGCCGCAACTACTCCTACCAGAACCTTAACCGCGTGATGATGTTCTATGATCAAAGCGTA
CAGGCCAACGGCTTTATTGGAATCGGTGGTTCCGGAACCACAATGGATGATCTGAACGGC
AATCAGCTGGACAATGCGCAGGCCGGGTTTGTGCGCGGCGGAATCATCTGGGCAAGGCAG
CCAGGCAATGGTCCGGTGCGCGGCGTTGCTGTGCCAAAAGGTACCCCGGGTTGGGGATCG
GCCTGGAAAAAGCAGTTTCAGAAAGCTTCCGTCATTCTGTTCTATTACGAAGTCCAGGGA
GCCTGCATGTCTTATCAGCAAACTATCTGAGCCTGGACCCTACCTGGAAAGATGCTTTT
GGCCGTCCTCTGTTACGGATGACTTTTGACTGGCAGCCTAACGAAGTTAAGGCTTCACAG
TTTCTGGTCGGGAAAGCGGTGGATATGTGTCAGGTACTCAATCCGAAATCTATCTCCAGC
GATGCCAAAAAAGACGGCGCTCATTACGATATCACCAAATATCAGAGTACCCACACCTGT
GGCGGTGCGGTAATGGGCAGCGACCCGAAAAAATCGGCACTGAACCGTTATCTGCAATCC
TGGGATGTGCCAAACGTGTTTGCTATCGGGGCGAATGCCTTTCCACAGAATAATGGATAC
AACCCTACCGGTCTGGTGGGCGGGCTGGCCTACTGGGCTGCCACCGCTATTCGTGAACAA
TACCTGAAAAACCCGGGCCCGTTGGTCCAGGCA

SEQ ID NO.64 - 3652

ATGGCCAGTGTAATGAAAAAACAGATGCAGTGATTGTCGGTTTCGGCTGGGCCGGGGCT
ATCATGGCAAAAGAGCTGACAGAAGCCGGGCTGAATGTTGTTGCGCTGGAACGTGGACCA
CAGCGTGATACTTATCCGGACGGTGCTTACCCGCGAGTCTATTGATGAGCTGACTTACAAC
ATCCGTAAGAAGCTGTTCCAGGATTTATCAAAAAGTACGGTCACTGTTCGCCACAATCCA
TCACAGACCGCCGTGCCTTATCGCCAGTTAAATGCGTTCTGCCTGGAACCGGTACCGGA

WO 2005/012547

PCT/US2004/024204

FIGURE 9F

GCGCCGGGCTGCACTGGTCAGGGGTTCAATTTTCGTGTTGATCCGGCTGAACTTCGCCTG
CGTAGCCACTATGAAGAGCGCTATGGTAAAGACTTTATCCCTCAGGGGATGACCATCCAG
GATTTTGGTGTGAGTTACGATGAACTGGAGCCATTCTTTGATCAGGCAGAAAAGGTCTTT
GGTACATCCGGAACCGCCTGGACAGTGAAAGGTGAGTTGGTCGGGAAAGGTAAAGGGGGC
AACCCGTTTGCACCTGATCGTTCCAGTGACTTTCCTCTGAAAGCACAGAAACGTACTTAT
TCAGCGCAATTGTTTGCAGAGGCTGCGGAGTCTGTAGGCTATCATCCGTATGATATGCCG
TCTGCTAACACTTCTGGTCCTTACACCAATACTTATGGTGCTCAGATGGGGCCGTGTAAT
TTCTGCGGATATTGCAGTGGTTATGCCTGCTATATGTACTCGAAGGCTTCGCCAAACGTC
AATATATTGCCAGCCTTACGTCAGGAACCTAAATTTGAATTGCGCAATGATTCCTATGTG
CTGCGGGTTAATCTGACCGATGATAAGAAACGCGCTACCGGAGTCACCTATGTCGATGCC
AGCGGTCGTGAGTTTGAACAGCCTGCTGACCTGGTCATTCTGTCTGCGTTCCAGTTTCAC
AATGTCCATCTGATGCTGCTTTCGGGTATCGGTAAACCCTATAACCCGGTGACTAATGAA
GGGGTTGTTGGGCGTAACCTTTGCTTACCAGAACATTTCTACTCTGAAGGCTTTGTTTCGAC
AAAAATATCACAACCAATCCGTTTATCGGTGCGGGTGGTGCGGGTGTGCGGGTGGATGAT
TTCAATGCCGATAACTTTGACCATGGTCCGCACGGTTTTGTGCGCGGATCGCCGTTATGG
GTAAACCAGGCCGGTGTTAAACCAATCTCCGGACTGCCGACTCCTACCGGAACCTCCGGCC
TGGGGTAGCGAATGGAAAGCCGCTGTCGCAGACACTTATACCCATCATGTATCGATGGAT
GCGCACGGTGCTCACCAGTCTTACCGTACTAACTATCTCGATCTTGACCCTAACTACAAG
GATGTCCACGGACAACCCTTGCTGCGCATGACTTTTGACTGGCAAGAGAACGACATCAAA
ATGTCGCAGTTTATGGTGTGCGAAAATGCACAACATCGCACAGGCCATGAACCCGAAAATG
ATCATGGGTGGGCCAAAGACCGCCGGTACGCATTTTGATACCACCGTGTATCAGACCACT
CATATGAATGGCGGCGCCATTATGGGTGAAGATCCGAAAACCAAGTGGATCAACCGCTAT
CTGCAGAGTTGGGATGTGTGCGAATGTGTTTGTACCAGGGGCTTCTGCGTTCCCGCAGGGG
TTGGGTTATAACCCGACCGGGATGGTCGCTGCACTGACCTATTGGTCAGCAAAAACCATC
CGGGAAGTTTACCTGAAAAATCCGGGTCCGTTGGTACAGGCA

SEQ ID NO. 65 - Orf 1952

ATGACAAAAAACTCCCTGCCACTGATGTGGTGATTGTTGGCCTCGGCTGGGCCGGTTCA
ATTTTAGCCAAAGAACTCTGCGATCAGGGTCTGAATGTGATTGGTCTGGAACGTGGCCCG
TGGCGCGATACTGCCAAAGATTTTAACGTAGCCACCGCGCCGGATGAATTGCGTTACAAC
GCCCCGCGAGGAACTGATGCTCCGGCCGGCACAGAACACCTGTACCATGCGTAATAACCCT
TCAGAACTGCACTGCCAATGCGGTTCTGGGGCTCTTTCCACCCGGGTAAACGGTACCGGT
GGCGCCGGTAACCACTGGGCGGGTATTACTTTCCGCTATCAGCCTGCCGATTTCCGGCTG
GCGAGCCATCTGCGTGAGCGCTACGGCAAAGAAGTTGATCCGGCTCTGACCTTACAGGAC
TGGGGGATCACCTGGGAAGAGATGGAACCCTTTTATGACTCTTTTGAACGGGTGCGCGA

WO 2005/012547

PCT/US2004/024204

FIGURE 9G

ATTCAGGGAAAGCCGGTAACATCAAAGGCAGTATTATCGAAGGCGGAAACCCGTTTCGAA
GGTCCGCGCGCCCGTGATTATCCAAATCCGCCAAATATTAGACCATTCGCGCAAACATTC
TTTGCCAAAACTGCCACCGAAATGGGTTACAAACCTTTTAACGTTCTTCTGCACTGGCC
TCTCAGGGTTATACCAACCAGTATGGTGTGACGATGGGCCCTTGTACCTATTGTGGTTTC
TGTACCAACTACGGCTGCGCCAACTATTCGAAAGCCAGTGCTATCGTTAACGTTTTACCG
GCAGTAGTCAGCATGCCCAACTTTGAAGCACGCACCAACTGTGAAGTCATGGAAGTTCTG
AAAGACAGCAGCGGCAAAAAGGCTACCGGAGTGGTGTATATCGACAGTAATGGCGAACGT
TATGAACAGCCAGCGTCTATCGTGATTGTGGCCGCCTTTACTTTTGAAAACGTGCGTCTG
ATGCTGCTGTCTGAATGTTGGTGTGCCTTATGACCCTGTCACCGGCAAAGGAACTACCGGC
CGTAACTACTGTTACCAGACAGCGAATGGTGTCCGTCTGTTCTTCAAAGATCAAATCTTC
AACCCGTTTATCGGCGGTGGCGCCATCGGGATGGGAATCGACGAATTCAACAACGACAAC
TTTGATCACTCCGGATTAGGGTTTGTGGCGGTGGTTCTACCCGGGTGACACCTATCGGT
GCTGCACCTATTGCTTCGCGTCCGGTGCCGCCGGGCACTCCGCGCTGGGGATCGGCCTGG
AAGAAAGCAACCGTGGAACATTATCTGACCAATATGAGTATTGGTTGCGAGGCCAGCAGC
TATCCGCAGCGCACCAACTATCTGTGCTCGATCCAACTATAACCGACCCGCATGGCCGG
CCGTTATTACGGATAACGTTTGATTTCCTGGACAATGATATGCGGATGGCACAGTATGTC
ACCAATAAAGTTGGTGAAATTGCCTTACGTATGAATCCGGTGCAAATCCAGAAACAACCA
CGTACTGCCCCGTGGGCCAACAACGACTACCAGTCATCTCACGTGGTTGGCGGCTTTGTG
ATGGGCGCTGACCCGTCTACCAGTGCGGTCAATAAGTTTTGCCAGGTCTGGGATATTCCA
AATCTGTTCTGTGGTGGTGGTCTGCGGTTCCAAATAACCCCGGGTATAACCCGACCGGT
ACTGTTGGCGCCCTTGCCTTCCGTACCGCTCACTATATCCGTACTCAGTACCTCAAACAG
CCCGGGGAGATGATGGTA

SEQ ID NO.66 - Orf 1749

ATGAGCAAATCAGGCCAAAGGCCGACGCAGTGATTGTGGGCCTGGGATGGGCAGGATCG
TTAATGGCAAATGAACTGACCCAGGCCGGGTAAATGTTGTTGCCATTGAAAGGGGGTCA
TGGCGTGACACCAGTACCGATTTCCCAACCTCAATTGATACTGACGAATTACGTTTCGTC
AGCCGCAGAGCAATAATGCAGCCCACTGCTGTAGAACTATGACCTTCAGGAATAATCCT
CTGCAGCAGGCTCTTCCACTGCGTGAGTTTAATACCTATCAGTTTGGTATGAATGTCGGC
GGTGCCGGTACACACTGGAACGCCATGACATGGCGCTTTTTACCTAATGACTTCCAGACA
TACACCAATACTGTTGAACGTTACGGTAAAAACAAATTCCTGGAAGGGATGCAGGTTTCAG
GACTGGGGCGTGACCTATGATGACCTTGAACCTTTTTATGATAAATTTGAGCGCTTTGCA
GGAACCTTCAGGAAAAGCCGGAAATATTAAAGGCGAAAAAATCGACGGCGGAAATGTATTT

FIGURE 9H

GAAGGACCGCGCAGCAGGGACTATCCGTTGCCACCACTGAAACGCACTCAGTTATCGATG
ATATTTGACAAAGCGACCCGCGAAATGGGGCTACACCCGTTTGCGGTACCTGCCGGAAAT
ACTTCAGGGGCTTATACCAATACTCTGGGCATTAATATGGCCCCTTGTAATTATTGTGGG
TTCTGTGAATTTTTTCGGTTGTGGTAACTGGTTCGAAAAGTAGTCCGAACGCCTGTATTTTG
CCAGCAGTGATGCAACGCAGTAATTTCTCAGTGATCACCGAGTCGGAAGTGTTACGGGTG
AATAAAGCGGCAGACGGCAAGACCGCCACCGGTGTCACTTTTATCGGAAGCGATGGTGTT
GAGTGGGAACAACCTGCCGATATCGTGATCATTTCTGCCTACCAGTTTGATAATGTCCGT
CTGATGTTACTGTCAGGAATTGGCGAGCCTTATAACTATAAAACCGGCACTGGTGTGGTT
GGCCGTAATTACGCTTATCAGACTATCTCTGGTGCAGGTGTATTCTTTGAAAATGAAAAT
CTGAATCCATTTATTGGTGCTGGTGCCTCAGGCAGTGGATGATTACAACAGCGAT
AACTTCGATCACTCAAACCTGGACTTTATCGGAGGCGGTGTGGCCCTGGTTCCTCCAGC
AATGGTTCGCCCCGATTGCATTATCCGGTGCGGTGCCTCCGGGAACACCAAAATGGGGTTCT
AAATGGAAGCAGGCTGCTCAGCAGAGTTACCAGAATTACAACCTCCGTCTATGTGATGGGC
AACAGCTATCCACATCGTGATGTTTTCTCGATCTCGACCCCGAATATAAAGACCGCCAC
GGACAACCGCTGCTCCGGGTACCTTCGACTGGATCGAAAATGATAAGCGCTCAGGCCAC
TTTATGGCCGATCGCTCGGTAGAAATTGGCCATGCGATGGGAGCAAAAACCGTAGTACGT
CAGGAACCTACGGCCCGTAATTTCTCGCCAATGGACAATCTTTCTTCGCATACCACCGGA
GGCGCCTGCATGGGAGATGATCCTAAGACCAGTGCGGTTAACCGTTACCTCCAAAGCTGG
GATGTGCATAACGTTTTTGTCTGTGGTGCCTCTGCTTTCGCGAATAATGGCGGTTACAAC
CCAACCGGTACCGTAGGCGCGTTAACTTTATGGGCTGCTGAAGCCATCAAAAATCAGTAC
CTCAAATCTCCCGGCCCGCTGGTGAGGATC

SEQ ID NO.67 - RZP 2633

ATGAATAATATTCGTCCAAAAGCTGATGTTGTTATTGTTCGGTCTGGGATGGTGCGGCTCT
TTAATTGCTGAAGAACTCACCCGGGCCGGAATGAATGTGGTGGCTATAGAACGTGGACCC
TGGTGGGAAACCAGCACCGACTTCCCGCCATCCATTGATACTGATGAGCTGCGCTGGGAC
ACCCGTCGTTCCATGTTGTTACCCCCGGCGGTTCGAAACGACAACCTTTCCGGAATAATACT
TCACAGCAAGCCTTACCTTCCCGTGACTGGAACCTGAATGAACTTGGCTATAACGTCGGT
GGATCCGGTACTCACTGGGCGGGAATGGCGTGGCGTTTTACGCCGTTTGATTTCCAGCCC
TATTCGCAAACCGTGGCGCGTTACGGTAAGCAGCAGATTGTTCCGGGGCTGATTCTGCAG
GACTGGGGGGTGAGTTATGATGAACTGGAGCCTTTCTATGATCGGTTTGAGAAAATAGCC
GGAGTTTCCGGAAAGGCCGGTAAGTCAAATGGCAACGTGGTGCCGGAAGGTAATCCGTTT
GAGGGTAACCGGAGCAGTGAGTATCCACTGCCGCCGCTGGAAAGTACCCGGCTAACCGAT

FIGURE 9I

CTGTTTGAGCAGGGGGCAAATCTCTCGGGCTCAATCCGTTTATGGTTCCTGCCGGTCAG
GCTTCCCGTGCCTATGTTAACCCGTTAGGGGTTCGCATGGGGCCGTGCACCTATTGTGGT
TATTGTTTGTACTATGGCTGCGGGAACCTTCTCTAAATCCAGCCCGAATGCCTGTGTTATT
CCCGCACTGATGCAGCGCGAAACTTTACTGTGCTTACTGATTCTGCCGTTGTGAAGGTT
AATAAGGCAGAGGACGGTAAGACGGCTACCGGTGTGACGTTTATTGATAAAAACAATAAG
CAGTGGGAGCAACCGGCAGACATTGTCATCCTGTCCGCATTCCAGATGCAGAATGTGCGG
TTGCTGTTGCTGTCTCAGATTGGTCAGCCTTATAACCCGCAGACAAAGCAAGGGGTGGTA
GGGCGGGCTTACAGCTTCCAGACTGTCTCAGGAGCCAGCCTGTTCTTCAAAGATGAGTAC
CTGAATCAGTACATCGGTGCCGGAGCGCTTTCACAGCAGGTTGATGACTTTAACGGTGAT
AATTTTGACCATACCGGTAAAGGGTTTATCGGCGGTGCCGGGATTCTGGTCGTTGCCCGT
GGAGCCAGGCCAATAGGTAATGCCGATACTCTGCCACCGGGAACCTCCTCGCTGGGGGAAA
GAATGGAAGCAGGCATATACCCATGCCTTCCAGAATGCGACCTTTATCTTTGGTCAGGGC
ACCAGTTATTCCCATGAAGATTATTATCTGGACCTGGATCCGGAATACAAAGATAAGTAC
GGATTGCCGTTACTTCGGGTGACGTTTGACTACAACGATAATGACCGGCGTTCGGCCAAA
TTTGTGCAACAGCGCAGTGTGGAAATCGGTAAAGCGATGGGAGCTGAGCGGGTTTTTGGC
ACTAACTCCGCTTCCGGCCATTACTCACCGTACAATTCGCCAGCGATCACACTATCGGC
GGAGCCGTCATGGGAACCGATCCCCGGACCAGTGTGTAAATCGCTACCAGCAGAGCTGG
GATGTGCACAATGTGTTGCTGCTGGGTGCTTCTTCTTTCCCGAATAATGCCGGATATAAC
CCGACAGGCACCATCGGTGCTCTTAGCTTATGGACTGCGAAAGCGATTATTGAGCAGTAC
AGAAAGAATCCTGGCCCGCTGGTGAAGGTG

SEQ ID NO.68 - Orf 3052

ATGAATTACACCAGACCTAAAGCAGATGCCGTTATTATCGGCCTCGGCTGGGCCGGCTCT
TTGATGGCCGAAGAACTGACCCGCGCCGGATTAAATGTGGTTGCGATTGAACGCGGCCCC
TGGGAACAGACCCAGAGCAATTTCTCCCCCGCCATTGCTGCCGATGAACTACGTTATGGC
GTACGCCGCGAAATACTGAAACCGCCCCGTGTAGAAACCCTGACTTTTCGTAACGACAGC
AGCCAGAAAGCACTGCCAGCCCGCGACTGGAATGCCTTTCAGATGGGCTACAGTGTAGGT
GGTGCCGGTAAACACTGGGCAGCCAATGCCTGGCGTTTTTAACCCTTCTGATTTTCGAGATG
GCGACCCGTCACAAAGAGCGCTATAACAATATGCCGCTGGCCGACGGCCTGATCCTTCAG
GACTGGGGGGTTAGCTATGCAGAGCTGGAACCTTTCTATGATCGTGTGGAAAAAATTGCC
GGTATTTCCGGTAAAGCAGGTGTGCTTAACGGCTCCACTCAGGAAGGCGGAAATCCGTTT
GAAGGCAACCGTACCAGTGAGTACCCGACACCGCCATTAATCCGTTTCGCACTGGAACGAT
ACCTTCCATAATATCACCACTAAAATGGGCTACCACCCGTTCCCTATTCCGGCAGGAACC

FIGURE 9J

ATCGGTGCAGCGTTTACCAACCCGTTAGGCATTAACCTGGCCCCTTGTACCTATTGCGGC
TATTGCGGTTTTTACGGATGTGGTAACTGGTCTAAATCTTCACCGAATATCTGTGTGGTT
CCGGCACTGATGGATCGCACCAACTTTACCCTGCTGACCGAATGTACCGCGCTGTACATT
AATAAAGCTGACGATGAAAAAACCGTCACCGGTGTGACCTTCCGTGATTCTGATGGTAAAC
ACCGGCTTCCAGCCGGCAGACATCGTTTGCCTGTGCGCGTATCAGTTAGACAACGTGCGC
CTGCTGTTACTGTGAAAATCGGTAAAGCCTACGACCACGCCACAGGCGAAGGTACACTC
GGTCGCGCCTATAACTACCAGACCATGTCGATGGGATATCTGTATTACGAAAATGAATAC
ATGAACCCGTTTCATCTCTACCGGCGCGTTATCCACCCAGATTGATGATTTCAATGGCGAT
AACTTCGACCACACCGGGCTTGGATTCTTGGGAGGTGCCGGAATTCAGGCGCTGTCAGAT
CAGGGGACTCCGTTAAGTATGACTGACCGCCTGCCTGCCGGCAGCAAAATGTGGGGCTCG
GCGTGGA AAAAAGCGTTCCGCCACAGTTATCAGAACTATGCCAAAATTCAGGGCCAGGGA
ACCTCGTATTCACACCGTGATTCCTATCTGTCACTGGACCCGAATTACACTGACGAGAAC
GGACAGCCTCTGCTGCGTCTGACCTTCGACTACAACCAGAATGATCGTCTGATGGCGCGT
TTTATTCGTGACCGGATTGAAGATATCTGTAAAGTCTCCGGCGCCAGCAGCTGGATAACC
GAAGCGTTCCCTGACTCCCACAACTACCGTTCCGTGCCTATGACAGTTCTCATACCATT
GGCGGTGCCGTGATGGGGCTTGATCCGAAAACCTCGGTACTGAATCGCTACCAGCAACAC
TGGGATGCACACAATCTGTTTGTGCTCGGGGCCTCATCTTACCCGAATAATGGCGGTTAT
AACCCGACCATCACTCTCAGTGCCCTGACGTTATGGACTGCGCACCATATCGTTAATGAC
TACCTGAAAAATCCGGGTTCGCTGGTACGA

WO 2005/012547

PCT/US2004/024204

FIGURE 10A

SEQ ID NO.69 - Orf 3818

ATGAGTCAGAATAACGTAGATGCAGAAGTAATTATTATTGGTTCCGGAGTGATGGGCGGG
CTACTGGCGACTCAGCTGTCTGCTGCCGGAAAATCAGTGATTATTGTTGAGGCGGGCCCA
CGTGTTACACGCCAGCAAATCGTTGACCGTTTCAGAACTCACCGTTCAAGATGTCACCTC
ACTAATATGAAATTGCAGGGTGTCTGGCTCGCCGTATCCTGATTTACCTCATGTACCTTCC
ACTTACGGCAACTACCTGCAACAGGTCTGGGCCGGTAAAATACCCGACTAAATACCTGCGT
GTCGTTGGCGGAACCATGGCACTTTGGTTCCGCTTTGTGGCGGATGATTCCAAACGAT
TTTAAATTGAAGACGTTGTACGGACATGGCAGAGACTGGCCGTTTGGTTACGACGAACTG
GAACCCTGGTATTGCGAAGCAGAGCATGCACTGGGAGTGTGAGGAGTCGATGGACAGGAC
GAGAGCGGGCACGGTGGTAAACCCTGGCCTCCACGCTCTAAACCTTTCCCGATGCCGGGC
TTACCTACCAGTTATATGTTTGACCGGCTGAGCGAGTTGCTTGGTAAAGGCGGTTATAAC
CCGGTGCTGGAACCAAATGGCAGGGCGACCCGCCCTGGGGCAATCGTCCGGTATGTGCA
GGCAATAATAACTGCAACCCGGTGTGCCCCGATTGGTGCCAAATATGATGGTTCAATGCAT
ATCGATCAGGCAGAACGGCTGGGTGCTAAGCTGCTGGACAATTCGGTGGTGTATAAAATA
GAAGCGGATGACAACGGTAAAATCACCCGTATCTGGTATAAAAAGCCTGATGGTTCTGAG
CATTCCTGACAGCAAATCTGTTTATTGTGGCTGCCTATGGGATTGAATCACCGAAATTA
TTGCTGATGTCTACCTCAGAAAAATATCCGAATGGTATTGCCAACTCTTCTGATCAGGTC
GGCCGTAATCTGATGGGTACACCCGGCATTAGTATGAACTTTATGATGGCCGAGGATGTC
TGGCCGGGGCAGGGACCAACAGAATTACTGGTCTATCTAAACAACCGTGATGGTGAGTTC
AGAAAAACGTTCCCGAGCTATAAAATCAAAGTGCGTAATACCGTGCCAACCGCTGATTAT
GCTTCCGGGCTGATTAGTAAAGGGGTATTAGGTTGAGAACTGGATGAGCAACTGCGTAAG
CTGTCAGCCCGTTGCTGAACTTCGCCATTGATTTTGAGACAGTGCCATTGCCGGAAGAAC
CGGGTCGTTCCAGTAAACCAAACTGATGCAATTGGAATTCCGCTACCGGAATATCG
TACAGCGTGACCGATTACTGGCAGGCAGGGAAAGAAGCAGGGCTGAAAGATTTTGCGAAT
TTTGCAAAATTAAGGGGGCGATGTGCTGAAAATCGACACCAATTATCAGGATCGCCAG
CATATTATGGGAACCATCATGGGCGATGACCCGAAAAATTCGGTGGTGAATTCTGAT
TGCCGGACCATGATCACCCGAACCTGTACATTGCCGGAACAGCGTGATGCCTTCGGCA
TCCTGTATGAATCCAACCTCTACCGGGGCTGCGTTAAGTTTGCGGCTGGCAAATCATCTG
TTGAAAAACGTACTGGTT

SEQ ID NO.70 - Orf 2448

ATGGCTAATACTACTTTAGATTTGACTACGTTATCGTCGGCTCAGGCGTTACCGGAGCA
CTGATTGCCTGGCAATTGTCCCGTCATGGTAAACGGGTTTGTATGATTGAGGCCGGCGAT

WO 2005/012547

PCT/US2004/024204

FIGURE 10B

CACATTCAACGCTGGAAAGCGATAGAGCATTACCGCAGTTTGCCGGATAAAAGCATCGCC
AATAATTCACCCTACCCTAATCTGGAGTGGGCACCCAATCCCATTGGCGGGCATTATCTG
GAACAAAAGGGCCCGGTAAATTATGCGACGACTTATATCAGAATGGTGGGGGGAACCTACC
TGGCACTGGGACTCTGCGACCTGGAGACTGTTACCGTCAGATTTTGAACCTAAAACCCGC
TATGGCGTAGGCCGCGACTGGCCGATTGGTTACGAAGTGCTGGAACCCTGGTATCAGAAA
GCTGAAGAGCAACTGGGGGTCAATGGCTGGGATACTGAAGATCAGAGTGGCCAGGGCAAA
GATCATTACCCTCCACGTTCAACAACCTATCCTACTCCGGGCCACCCGTTTCAGTTGGGGA
CAACAGGTTGTTGCCGGTAAACTTGAGGCGGCGGGTTACTCTGCCATTCATGAACCCAAT
GCCAGGCTTTCTGTGGCTACCGCCGAACGTCCGGCCTGTGCCGGAAATAATACCTGTGAC
CCTATTTGCCCCATCGGCGCTAAATATAACCGCTGACTTTCATGTGCAAAAAGCGCTCGAC
CATGGCTGTACATTGCTTTCTAACTCGGTGGTTTACCGTGTGAGGCCGGAGATGATGGT
AAAATTACCGCTGTGCATTTCCGTCCGCCGGATAAGTCAACGGGAACAGTTAGCGGGAAA
GTTTTTGTGATTGCTGCTAATGCAATTGAGACGCCTAAATTGCTGCTGATGTCCGTATCG
GAGCGCTATCCTCAGGGTATTGCTAATACATCAGGTCAGGTCGGGCGAAATCTGATGGAT
CACACCGGACTGGGCTTTAATCTGGTGACCGAAGATGAAGTCTGGCCAGGTACCGGGCCT
AATGCCTTGCTGGTGATGCTCAATGCCCGTGAAGGGAAATTCAGAGCCGAAAGAGCGTCT
TATAAACTAAATTTCTGTAATACCGCCGTCAATTTTGCCGTCACCTAAGTCCCTGATTAAA
CAGGGAATCATGGGTAATGAGCTCTACCGGCAAATTAAATATCAGTCTGCCCCGTGAGTTA
TCTATTGCCGTCGATCTGGAAACTTTGCCTAACCCGCAGAATCGTATTGTCCCAAGTAAA
GATCGGACTGACAGCCTGGGGATCCCTGTCCCGGAAATCCACTACGATGTGGATGATTAC
TGGAATAAAGGGCGCGATGCTGCTATCGCTGACGTACAAAATATCGCAAAAATTCTGAAC
GCAAAGATTGTGGCAACGGATACCAATAAACAACCGCGAACATATTTTGGAACCATG
ATTATGGGTAACTCACCCACTGATTCGGTGGTGGATAAAAATTGCCGGACTCACGATCAT
CCCAATTTATATATCGCCGGGACCAGTGTCTTTCCGGCCGTCGGCTGTGTTAATCCGACC
CTGACAGGTGCAGCATTAGCTTTACGCATTGCCGACACGTTGTTACAGGATCCAGTGACT

SEQ ID NO.71 - Orf 3042

ATGAAAACCACACATTCTGCCACCGTTGTTATCATCGGTTTCAGGCATCGCAGGCTCACAG
ATAGCCCAAAAATTACAAAAGCTGGCATTGATACTCTGATGCTGGAAGCAGGTTTCACGG
ATTGAGCGCTGGAAAATTGTAGAAAACCTACCGTAATTCGCCATTTAAACTGATTTTCAG
TCACCGTATCCACCGACCCGTCATGCGCCACACCCGCAATACTCGCCGGAAGATAACGGC
TACTTTATTTCAGTATGGTCCGGAGCCTTATAAAGCGGGCTACCTGCGTGTGCTGGTGGGA
ACCACCTGGCACTGGTCTGCTCAGGCCTGGCGTTTACTGCCAAATGATATGCGTCTGAAA
ACACTGTATGGCGTCGGGCGTGACTGGCCGATCAGTTATGACGATCTGGAACCGTATTAC

WO 2005/012547

PCT/US2004/024204

FIGURE 10C

TATGAATCTGAAGTTGAAATGGGTGTTGGCGGCCCGGAAGACACCGGTTACCCACGGAGT
AAACCTTATCCACATCCGCCGTTGCCTTTGTCTGATTTTGACAAGGCCTTCAAAAATGTG
GTGGATAAAAACGGCTATCACCTGATTACCGAACCGGCAGCCCGTAATACTGAACCGTTT
GACGGTCGCCCCGGCCTGCTGTGGCAACAGTAACTGCATGCCTATTTGTCCTATTGAAGCG
CAGTACACCGGTGAAACCGCGGTTTCGTAAAGCTGAACGTGCCGGATCTCTGCTGGTTCCG
GATGCCGTGGTCTATAAAATAGAGCACGATGCCAAAGGCAATATCACTTCTGTTCTGTAT
AAAGATCCAAACGGGGAAAGTTTCCGTGTACAGGGAAAATCTTCGTA CTGGCGGCTAAT
GCCATCGAAACTCCGAAGCTGATGCTGATTTCCCGATCTGACAAATATCCAAATGGTATT
GGCAACACGACTGATAACGTTGGCCGTCACCTGATGGATCATCCTGGAACCTTCAGTTTAC
TTCCTGAGTAAAGAACCGATGTGGCCGGGCGTGGGCCTATGCGTTTAAGTTGCATTAAAC
AACCTGCGCGATGGTGATTTCCGTTCAGAACATTTCGGCGATGAAAATCAACCTGGGCAAC
TATTCTCCAACGCTGGCGGTGAGTAATTACCTGCTGAGTAAAGGGGTTTCCGGCAAAGAT
CTGCCGGCGATGGTGCGTGATTATGCGTCACGCTGGGTAGCAGTGAATACCTTCTTCGAT
ATCCTGCCTGACAGAGATAACCGGATTGTGGCGGTAGACTCGCAAAAAGATGCAATGGGT
ATTCCTAAACCAGGCGTGCATTACCACATCAACGATTACATTAATAAAGCGCGTGATGTC
GCCCACCAGCACTTCGACCATATTGCCGGTCTGTTTGGCGGTACGGAAGTGCGTCATGAT
GATAAGTACTTCAACAATAACCACATCATGGGCACCCTAATTATGGGTAATGATCCCAAC
GACTCCGTGGTGGATGCTGATTTACGTACCCACGATCACCAGAATCTGTTTGTGGCGTCC
AGTGGTGTTATGGCCAGTGCCGGAACCGTTAACTGCACCCTGACATTATCAGCGCTGGCG
ATGAGACTGGCAGACAAATTAATCGCGGAGTGCCAACATTTA

SEQ ID NO.72 - Orf 1983

TTGGCTACAGAATTTGATGCAGATGTGATTGTCGTAGGGTCAGGAGCCTGTGGCTCTAAC
CTGGCAAACGAACTGGCGGTAAAAGGGAAATCGGTGATTCTGCTGGAAGCAGGTGCCAAT
GTGCCGCGCTGGAAAATCCTCGAAAACCTCCGTAAATTCAGGCCGCCATTATGACCGGAAT
AATGCATACCCAAATAATCCGTGGTCTCCGACCAGTAATAACCCCGGCTATATAGAGAAC
GTTGGTGAATTTTCGTGAACAGCCTGGCATGCTAAAACCTGGTGGGTGGCACCACATGGCAC
TGGGGTGGCGCCACCTGGCGGTACATTCCAAATGATTTTAAACTGAAGACTATGTACGGT
GTGGGTGCGGACTGGCCGATCAGTTACAGCGATCTGGAACCCTTCTATACCCGTGCCGAA
TATGCTATCGGGGTGCGCGCTCAGACACAGAAGACCAGTCAGGTCAGAACCCGGGGATC
TCTTTTCCTCCACGTTCCAAAGCTTACCCTGTAGATCCGGAAGCTGATATCTACAGTAAC
GCCAAACTTAAAGCAGCACTGTTACCGCATGGCCATAGCGTAGTACACGAACCCACAGTG
CGTATCCATCGCCCTTATGATGGTTCGTCCGGGATGCCAGGGGAATAATAACTGTGACCAG

WO 2005/012547

PCT/US2004/024204

FIGURE 10D

GTCTGTCCAATAGGTACCTTGTATAACGGTTCAGTTCATGCAGATAAAGCCGTGCGTAAC
GGCGCAAACTGATTACCGATGCGGTAGTACACAAAATTACCAAGGGTGAACAAGGTAAA
ATCACTTCTGTCAGTTACCTGACCCCTGCTGGTGAAGAACACACGCTGACTGCTAAATAT
TTTGTTCTGGCGGCACACAGTTTCGAAACTTCAAAGCTGATGCTGATGAACGATATCGGT
AACTCCTCTGATATGGTGGGTCGTAATCTGATGGACCATATTGGCCTGAGTATGAATTC
CTTGCCGATGAACCAATGTGGGCAGGACGCGGCCCGGTGCAGCAGGCGACTATTATGACC
TGGCGTGACGGTGATTTTCGTTCCAAATATTAGCCAACAAGCACTCTTTAGCCAATAAC
AACCACAAATCGATATCGCCCAACGTGCTATTAACGAGGGATTGATGGGTAAAGAGCTG
GATGCTCGTATCCTTGACTGGTCATCCCGCTGGATGTCGATCTATAGCTTTCTGGAACCA
TTGCCTAATCCGGCTAACCGTGTACAGCCGAACCCGGCCTGGAAAGACAGCCTTGGTTTA
CCTGGTATCAAAGTGACCTTTGATGTCGATGACTATAACCAAAGTGGGTGCTAAGCACATG
GTCGAACAATATAAGCAGATTGCCGGGCTGATGAACGGTCAAATCATTGATTTAAATACT
GCGTTCGAAAACCATGACCACCTGATGGGCACCATGATTATGGGTGATAATCCTAAGGAC
TCCGTGGTTAACCATGAATGCCGCAGCCACGACCATCCGAACCTGTTTATTGCATCAGTT
GGCGTAATCCCTGCTGCCGGTGTGCTTAACCCGACACTGACCGGTGTGGCTCTGGCAATC
CGTTCTGCTGACATCATTGCAAAGGAGGTG

SEQ ID NO.73 - Orf 3399

ATGTCTGATTCATTAAGCGCCGACGTTGTTGTTATCGGAGCCGGGATTGCCGGTTCCTG
GCAGCACTAAAAATGGCTAAAGCGGGTGCTCTGTCTTAATCCTTGAATCCGGTCCGGAA
ATTAAGCGTGACGAAGCGGTGAACTATTTTCGCAACTCGCCATTTAAAGGCGATTTTACC
GAACCGTATCCGCCGGAACCCTGGGCACCTCAGCCTAAATTTATTCCGACTGATAACAAT
TATCTGATCCAGAAAGGTCCGGACCCTTATCGTGCCCAAGTATTTACGTGGCATCGGTGGT
ACCACCTGGCACTGGGCCGGTCAGGCGTTTCGTCTGCTGCCAAACGATATGAAGATCAAT
ACCCTGTACGGTGTTGGTCTGACTGGCCAATCAGCTATGAGGACTTAGAGCCTTACTAC
AGCGATGCGGAATACCAGATGGGAGTTTCCGGTGATGATGATTTAAATTCGCCGCGCTCA
CGTCCTTACCCGTTACCAGGCATTCCGTTACCTTACGGTTTTGAGCGCCTGAAACAGCGC
CTTAGCCCGCTTGGCTATCAGGTGGGTATCGGTCCGCAGGCACGTAAACAGTATTCCTTAT
CAGGGACGCCCCGGCCTGCTGTGGTAATAACAACGTATGCCTGTTTGTCTTATCGATGCT
CAGTACCACGGCGGGATCTCTGCCCGCAAAGCAGTGGATGCCGGGGTAAAAATTATCGCC
AATGCCGTGGTTTACCGCATCGAAGCAGATGATCATGGCGTGATCCAGGCTGTACATTAT
CTGGATCAAAACAAAGCGACTCACCGCGTGACCGGTAAACAGTTTGTCTGACCGCAAAC
GGCGTTGAAAGCCCGAAAATTCTGCTGCTGTGACCTCAGATCGCTATCCAAACGGTATT

WO 2005/012547

PCT/US2004/024204

FIGURE 10E

GCTAACAGCTCCGGGATGGTAGGACGTAACCTGATGGACCACCCGGGCACCTCCGTGAG
TTTTATGCTGACGAGCCAATCTGGTTTGGTCGTGGTCCGATGCGTCCGGGCAGTATCAAC
AACATGCGTGACGGTAGCTGGCGTAGCGAGCGTTCCGCATTGCGTATCGACCTGGCTAAC
ACCTCGCCGGTGCGTTATCTGACCGAGCGTCTGGTACGTCAGGGTTATTACGGCAAAGCG
CTGAACGACAAACTGGCCTTCCAGGCCGAGCGTTTTGTACAGCTGAAATGCCTGCTGGAA
ATGCTGCCGGATCCGGAAAACCGCTTAGTCCTCAGCAAAACTGAGAAAGATGCCTGGGGT
ATCCCGCGTCTTGAGGTGTATTACAAATTCCTGAATACGTGCATGCCGGTTATGACCAG
TCTATGTCTGACTTCCGGAAAATTGTTTCAGCAGATGGGTGGAACCGAGCCGCTATATAGC
CAGCGTGGTGTCTACGACAACAACCAGCATATCACCGGCACCATGATTATGGGCAGCGAT
CCTAAGAACTCCGTCGTTGACGGTAACTGTCGTACCCATGACCATCCGAACCTGTTTATT
GCCGGTACCGGAATCATGCCTTCGGCGTCAACCGTTAACTCCACTTTAACGGGGACGGCT
CTGGCGTTGCGTATGGCCGACTATGTGCTGAAAAGCCTG

SEQ ID NO.74 - Orf 1955

ATGTCTGAACAATATAGCGCAGATGTGGTAGTTGTTGGTGGCGGTATTTGTGGTGGAACC
GTGGCAAAAGAGCTGGCAGAAGCTGGTTTATCCGTACTGGTACTGGATGCAGGCCCTCGC
TGGGAAAGAGGGGAAGTGGTAGAGAACTGGCGAAATCTGCCGCCAGTGAATAAATCTGAA
TCTGATTACGCAACGCCATATCCGGCGGAACCGTGGGCAGTTCATCCGCAGCTCTATCCA
TATAATAATTATCCGGAAGTTTCCGGTCCGGATGCTTCTGCATTTTCGTCAGGGTATGATC
AAAGGCGTCGGCGGGACAACCTGGCACTGGGCAGCTTCCTGCTGGCGTTTCCTGCCTGCA
GATATGCAGCTCCAGACCACTTATGGTGTGGGCCGTGACTGGGTGGTGA CTTATGACGAA
ATGGAAGATTATTACTATCGTGCCGAGGTGCTGATTGGTGTTAATGGCCCTAACGACACA
TCACTGAAGTACGTTGCTCCACGTAAAAAGCCATTCCCGATGGAGCCAATGCCATACGGT
CCGGCCGATCGTCGGTTTACCGAAGTCGTGGCAACTGCCGGTTATGAAAACACCCCGGTA
CCCCAGGGACGTAAACAGCCGGCCATATGACGGGCGTCCGCAGTGCTGCGGTAACAATAAC
TGTATGCCTATCTGTCCAATCGGCGCCATGTTCAATGGTATTCACAGTATTATTAAAGCC
GAAAAAGCAGGTGCTAAGGTTCTGCCAAATGCGGTGGTCTATAAGTTTGATACCGACGAA
AACAACAACATTACAGCGTTGTACTACTACGATCCGGACAAAAACTCGCACCGCGTAACA
GCACGGACTTTTGTGCTGGCAGGTAACGGTATCGAAACACCGAAACTGCTGCTAATGGCG
GCTAATGACCGTAACCCGAATGGTATCGCGAACAGTTCAGGAATGGTGGGGCGGAATATG
ATGGATCACCCGGGTATTCTGATGAGCTTCCAGTCAGCAGAGCCTATCTGGACCGGCGGT
GGTTCGGTACAGATGAGTTCAATCACTAACTATCGTGATGGTGATTTCCGTCGTGAGCAT
TCGGCAATTCAGATCGGGATGAACAATACTTCTCAGAACCATAAAGCCGGGGTGAAAGCG
CTGCAAATGGGGCTGGTAGGTAAGAACTGGATGAGGAGATTTCGTCGCCGGGCCGCTGT

WO 2005/012547

PCT/US2004/024204

FIGURE 10F

GGAATGGATATCTATGTTAACCACGATATTCTGGCCAATCCGGACAACCGTCTGACACTC
AGCACCGTGCATAAAGATAAACTGGGTATCCCATACCCGCATGTTACCTATGATGTGGGT
GATTATGTGCGTAAGGCCGCGGTTTCATCCCGTGAGCATCTGATGACTATTGCCAAACTG
TTTGGCGCGACCGAAATCGAAATGACTCCGTATTTTAACCCGAACAACCACATTATGGGC
GGGACTATCGGGGGACATGATCCGAAAGATTGAGTGGTCGATAAATGGATGCGAACCCAC
GATCATCAGAACCTGTATATCGCTTCCGGTGGTGTGATGGCAGCCGCAGGGACTGTTAAC
TCAACCCTGAGTATGGTGGCGTTATCTTTGCGCGCTACTGACAGTATCAAACGCGATCTG
CAACACGGT

SEQ ID NO.75 - Orf 3675

ATGAATGCAGATGTGATTGTGGTTGGCACCGGGGTTGTGCGTTGCCTGATTGCTGAACAG
TACTCGATAGTGGCCACTCCGTTGTGATGCTGGAAGCTGGCCCGCGGGTGAACGTTGG
CAGATTGTGAAAACCTACCGAAATTTACCACCGGTTTCCCGGTTACATTTTAATGCCCT
TATCCACCGGAGCCCTGGGCTCCTCATCTGATGTCGGCCACGCCTGAGCAGGCCGCTGAA
TATCTGCAACTGGAAGGTCCTAATGCCCGGGCCTATCAGCAGGGATATGTTGTTATGCC
GGCGGAGCAACATGGCACTGGGCTGGAATTTGTTGGCGGTTAACTCCGGAAGATATGCAG
CTAAAACGCTCTACGGTGTGCGCGTGACTGGGCTTTTGATTATGCCACCCTGGAACCT
TATTACACCCGTGCCGAATATGCTCTTGGGGTCTGTGGACCTTCTGAGCCGGAACCTGCAA
TGGCCCCCGGTTGTTCCAAACCTTATCCGATGGGCGCTCTGCCTTTCGGGCCGGGTGAG
CAACGCTTTACCGACGCGGCTGCCTCTATTGGCCTGACAAATCTGCCTTCAGCTCAGGCG
CGTAATAGCGGTATTGCTTACGGGGATCGTCCTGCCTGTTGCGGTAATAATAACTGCATT
CCGGTCTGCCCCGATTGGTGCGAAGTACGATGCAGCAACTTCACTGACACGTATTGAATCC
AAAGGCGGCAAAATTCAGCCAAATGCGGTGGTCTATAAAATAGAAACCGGTGCGGATAAT
AAAGTCCAGGCAGTTCCTATTTTGACAATAACAAACAGACTCACCGGGTGACCGGGTCT
GTCTTTGTTATTGCCTGTAACGGTATTGAAACTCCGAAACTACTACTGATGTCTGCGGAC
AGCAGAAACCCGCACGGTGTAGCAAACAGCTCAGATCAGGTGGGACGCAATATGATGGAC
CAACCGAAGCTGGTAGTAGAACTTGAGCTGGCTGAACCCGCCTGGACCGGCGTTGGACCG
GTGCAGGGAAGCAGTATTATGGAACCTCTCAGGGCAGTTTCCGTTCTGAATATTGCGGG
GCGTTGTTCCGTTTCAACAATATGGCTCGCAGCCGGATTGGTGCTATGGCGGCGCTGGAA
AAAGGTCTGGTGGGTAAAGCACTGGATACTGAAATTCGTCGTCTGTCAGCCTGTACCACC
GAAATTGCTATTGAACATGAGCTGATGCCGGATGCCAATAACCGCTTAACCTTTATCGGCT
AAAAAAGACTGGCTGGGATTACCTAAACCCAACATTTACTACGATGTCGGTGACTATGTA
CGCCAGGGATCGCAGCGCCATTCTTTACCTATCGCCCGCCAACCTGGCAAAGCGATGGGA
GCGACCAAGTAGATATTTGACCGAATATACCAACAGTGACCATATTATGGGCGGCTGC

WO 2005/012547

FIGURE 10G

ATTATGGGGACCGATCCTGCAGTCTCGGTGGTGGATGTGGATTGCCGGGCTCATGACCAT
GAAAACCTGTTTTTACCGGGTGGAGCTGCGATGACTACCGGAGGATGTGGTAACAGTACC
CTGACAATGTCTGCCTTAGCACTGAAAGCCGCAGATGCCATTTCATGCACAATTAGGGAAA
GCA

SEQ ID NO.76 - Orf 1220

ATGTCTGAGACTATTTCTACTGACATCGTCGTGATTGGTTCCGGCGTTGTAGGTTTCGTTA
ACAGCCAGAAAACCTGGCTTTGGCAGGACGTAAAGTTCTGATGCTGGAAGCAGGTCCGCGT
ATTCAGCGTGACCAGATTGTCAGCAATTTTCGCCATTCGGCACGTAAAGACGATTTTATC
GCCCCGTACCCAAACTCTGAAATCGCACCATTTCCTGACTATAAGCCGGAAGATAACGGC
TATTTAGACCAGACAGGGCCTAAAGATTACAAGCCGGAATATTTGCGTGTGTGTCGGTGGC
ACCAGCTGGCACTGGGCAGCTCAGGCCTGGCGCCTGGTTCCGAATGATTTTCGACTGAAA
TCACAGTATGGCGTGGGTGCTGACTGGCCGATCAGTTATGAAGATCTGGAACCCTATTAT
TATGAAGCCGAGATTTTGTGGGGTGTTCCTGGTCTGCAGAAATGGCTAAATACTCGCCG
CGCAAGCATCCCTATCCGATGGAAGGGGTGAAAATGTCTTATCTTGAACAACGGGTCACC
GCGCGACTGGCACCCAAATATGAAGTTCTGACCAACACCACCGGTGCTAACTCCGTTCT
TATGACGGTCGTCCACAGTGTTGCGGCAATAATACTGTATGCCAATCTGCCCCGATTGAT
GCACAGTATCACGGCGGTATTGCTGCTGCTGCCGCTGAGATTGCCGGGGTTAAACTGATC
CCTCAGGCAGTTGTTTATAAGCTTGAACATAACAGTCACGGTAAAATTACTGCTCTGCAC
TACTATGACTGGAATAACAATCGCATCGTGTGGAAGCTGAAATCTTCGTGATGGCGGCT
AACGCTGTCGAAACCCACGTATTCTGATGCTGTCAGCCGATGATAAAAACCCGAATGGG
TTATGCAATAACTATGATCAGCTGGGACGTAATCTGATGGATCACCCGTCGAATTCCGCA
ACTTTCTACGTAGACGAGCCTCTCTGGCCAGGTCGTGGACCGATGAGCCCTTCATCTATC
CAGCAATTGCGTGATGGTGCATTCCGTTTCAGAGTCAGCGGCTTTCGGTATTGATATCTCT
AACTCCTCACGGGTTGCCGGTGTTACTGCCGGAGCGATTAAAGAAGGCCTGACCGGAGCC
GATCTGGACAGCGCTATTCTGTATCGCGCCTCATATGAAGTGAAGCATTAAAAACGTTCTG
GAGCAGTTACCTGATCCGAAAAACCGCACCATGCTGAGCACCCGTAAAAAAGATGCCCTC
GGTCTGCCGGTTCCTGCGTTCTCGTACTCTTTTGATGAATATATCGAGAAAGGCATGCAG
CACTCGCTGGAAGTCTATGCCGATATCGCCCGCATGCTGGGTGCCACGAATGTCCGTTAT
TCGACTCCGGGTGTGTATAGCAACAACCAACACATCACCGGAACCCTGGCAATGGGCACC
GATGAAAAAACCTCAGTGACCGACCATGTGGGTAAAGCCTGGGAATACGACAACCTGTAT
ATGGTGTCTACCGGGGTGATGCCAACAGTGGCCACTGCTAACTCCACGCTGACCGCCTGT
GCTCTGGGCTTACGCACCGCTGACGCCATTCTTGGCAAAATC

WO 2005/012547

PCT/US2004/024204

FIGURE 10H

SEQ ID NO.77 - Orf 2419

ATGATGATGAAAAAACCAGAATTTACTCCGGGTGGCGATGCCTCCGCGGATATTGTTATT
GTGGGCTCCGGTATTGTTGGTGGACTGATTGCAGACAGACTGGTCAGTCAGGGATATTCC
GTACTGATACTTGAAGCAGGGTTACGAATCAGCCGTGCACAGGCAGTAGAAAACTGGCGT
AATATGCCGTTTGCTAACCGTGCCGGTTCAGATTTTCAGGGCTTATATCCGCAGTCACCA
CTGGCGCCTGCCCCGCTCTATTTTCCGCCGAACAACCTATGTCAATGTCACCGGACCAAGC
GCCGGCAGCTTCCAGCAAGGCTATCTGCGAACTGTCGGAGGCACCACCTGGCACTGGGCG
GCTTCCTGCTGGCGCCACCATCCAAGTGACTTTGTGATGAAAAGCAAATACGGTGTCCGC
CGCGACTGGCCTATCTCTTATGACGAGATGGAGCCATGGTATTGTGAAGCCGAATATGAA
ATTGGTGTGGCCGGCCCCGAGCGACCCGTCCATGCAGTCACCGAGTGAACGTAGCCGTCCT
TATCCGATGGATATGGTGCCATTTGCTCACGGTGATACTTATTTTGCCAGCGTGGTTAAC
CCGCATGGTTATAACCTGGTGCCAATCCCGCAGGGTCGTAGTACCCGTCCGTGGGAAGGA
CGCCCGGTTTGCTGCGGTAACAATAACTGCCAGCCTATCTGCCCAATCGGTGCAATGTAT
AACGGTATCCACCATATAGAGCGTGCTGAAAGCAAAGGTGCGGTGGTTCTGGCAGAATCA
GTGGTCTACAAGATTGATACTGATGATAATAACCGTGTTACTGCGGTGCACTGGCTGGAC
AACCAGGGCGCATCACACAAAGCGACCGGTAAAGCGTTCGCACTGGCCTGTAACGGGATT
GAAACCCCGCGTCTGCTATTACAAGCAGCCAATAAGGCTAACCCGACCGGGATTGCCAAC
AGCTCAGACATGGTTGGCCGTAACATGATGGACCACTCCGGCTTCCATTGCAGCTTCCTG
ACCGAAGAGCCTGTGTGGCTGGGTGCTGGCCCCGGCTCAGAGTAGCTGTATGGTCCGCCCG
CGTGACGGTGCCTTCCGTAGCGAATATTCGGCTAACAAAATGATCCTGAATAATATTTCA
CGGGTTGTTCCAGCCACCAACAGGCTCTGGCTAAAGGACTGGTCGGCAAAGCTCTGGAC
GAAGAGATTGTTATCGTTCTATTTCATGGTGTGATCTTTCCATCAGTCTGGAACCGTTA
CCAGACCCGGAAAACCGTCTGACTCTCAGCAAGACTCGTAAAGATCCACATGGCCTGGCC
TGTCGGGATATTCATTACGACGTGGGAGATTATGTGCGTAAAGGGGCGACTGCGGCTCAT
GAACAACCTGCAACACATCGGTTCTCTGTTTAATGGTAAAGAGTTCAATATCACGACTGCC
CTGAACGCCAATAACCACATTATGGGCGGAACCATCATGGGTAAAAGCGCCAAAGATGCC
GTGGTCGATGGTAACTGCCGGACCTTTGACCATGAGAATTTATGGTTGCCTGGCGGCGGA
GCCATTCTTCAGCCAGTGTGGTGAACAGTACTCTGAGCATGGCAGCACTGGGCCTGAAA
GCTGCACACGATATTTCTCTGCGCATGAAGGAGTTTCGCA

FIGURE 11A

SEQ-ID NO.78 - Orf 766

ATGAAAAAGATGACATTTAAGCGCCTGTTACTGGCGAATACTGTAGTTCTGGCCTGCGGG
CTGGCTGGCGCGGTACAGGCGGCCGATGCACCGAATCAGGATCAACTGGTAAAACAGGGT
GAATATCTGGCACGGCTGGGAGATTGTATGGCTTGCCATACGACCTCCGGGCGGCCTGAT
TATTCGGGCGGTCTGGCGATAAAATCGGATCTCGGGACTATTTACTCCACCAATATCACC
CCGGACAAACAGTACGGGATAGGTAATTACACCGAGCAACAGTTTGCCGATGCAGTGCGC
AAAGGGGTGCGTCCGGATGGCAGTTTCCTCTATCCGGCCATGCCTTATCCGGATTATGCC
AAAACCAGTGATGCGGATATTCATGCCCTGTACAGCTACTTTATGCACGGTGTGACCGCC
AGCAACAGTCAGCCGCCGCAGACCGACCTCAGCTTCCCGTTCAGTCAGCGTTGGGGCATG
CGTTTCTGGAACATGGTGTTTACCTCCGATAAGCCATTCCAGCCGATTGGCGGAGCTTCA
GAGCAGGTTAACCGTGGGGCTTATATTGTTGAGTCTCTGGGCCACTGTAGCAGCTGCCAT
ACGCCGCGTGGTGTGGCAATGGAAGAGAAAGCGCTGGACAGCAGTGACAGCAACTTCCTT
TCTGGCGGCAACCTCAATGGCTGGGATGTCCCTTCATTACGGGGTATTGCCCGCTGGAGC
CCGGATGAGATTGTCGATTACCTGCAAAGCGGACGTAACGACAAGGCCGGTGTGCGCCGGT
GAGATGACCTCGGTCGTGAAAAATTCGACCTCGCACATGACCGATGCCGACCTGCAGGCG
ATTGCCGCCTATCTGAAATTCCTTGGTGGCAACCCGCCATTGCAGGCTTATGATCAGCAG
AAAAATCAGGCCACTACCGCTAAACTGACCGCGGCGGTGGATCTGACTGAAGGCCAGACG
CTCTACCTGAACAACTGTGGTGCCTGCCATTTTGTTAACGGGCTGGATGCTGCACGGGCA
TTCCACAGTTGGATCAGGCATCGGTTGTTAATGCCAAAGATCCGCAGGGGCTGATCCAT
ATCATCCTGCAGGGTGCGCAGTTACCGGCAACTGAGAAGTCGCCATCAATGCTGAAAATG
CCTGGCTTCGGACACCGTTTATCTGACGACCAGGTGGCTAAACTGGCAACCTTCGTACGA
CAGGGCTGGAGTAACGATGCATCGGCAGTGACTGCTGATCAGGTGAAAAAAGTTCGTGAG
GGGCTGGAGCAGCAC

SEQ ID NO.79 - Orf 2908

ATGAAAACAATTTTTGTGAAACTTCTGCCTCTTGCCATAATGTCAGTTATTGGCGTTATC
GGGCTGAAGCAGGCTTATGCTGACAGCAATGATAGTGCAGACCTGATAAAACAAGGTGCA
TACCTGGCTCGCGCCGGAGACTGTACAGCCTGCCATACTGAAGCTGGCGGCAAACCCTTT
GCCGGTGGTCTGGCTATCAGGAGTCCGATGGGAGTCATTTACTCAACTAATATCACTCCG
GATAAAAATGCCGGAATCGGCAGTTACACCGAACAACAGTTTGCAGAGGCGGTTCGTAAA
GGAGTCCGTCGGGATGGCAGTAATCTGTACCCGGCAATGCCTTATCCTGACTATAGCGGT
ATTACCGATAAAGACATTCATGCCCTGTATGTGTACTTTATGCACGGTGTAGCCCCGGTG

FIGURE 11B

AGTGTAAGCACCACAAACCTCCCTGACTTTCCCGTTCAGCCTGCGGTGGGGAATGAAA
TTCTGGAATATTGCCTTCGCGTCCGGAAACAGCTATCCACCAGCTCCAACAACCTCAGTCA
GACAGTGCTGATGCTCAGGCATTAAGCCGGGGCAGATATCTGGTCGATACTTTAGGTCAC
TGTAGCAGTTGTCATACTCCACGAGGTATCGGGATGCAGGAAAAAGCGTTGAACGACAGT
GATAGTCGCTTTCTGTCTAGCGGCATGCTTAATGACTGGACAGTGCCTTCGTTGAGAAAT
CCTGACGGATGGTCTGTGAATGATATTGCAGAATACCTGTCTACAGGGCGCAATGACTTC
GCCAGTGTCGGTGGTGAAATGACGGGCGTGGTGCAACACAGCATGCAACATATGAACCAG
GCCGATTTACATGCCATTGCTCTGTACCTTAAATCATTACCTGCCAGTACTAAACAGCAG
CATAATGTGAAACCCGATCTGCAGAATGACACTCAGAAAACGGTGGATACTCTGACGCTC
GGCAAAAATCTCAACTCTGGTCAGATGCTTTACCTGAACAACTGTGAAGCATGTCACCTG
ACCGATGGCGGAGGAGCTAAAAAGATTTTCCACGTCTGAATGGGGCCAGTATAGTGCTT
GCTGATAACCCGACAGGGCTGATATCGGTGATGCTTAAAGGTGCGCAGACCCCTTCTACG
GCAAATGCACCGTCAGTACAGTTTATGCCCGGATTTGAGCAACGGCTCAATGATCAGCAA
ATTGCTGAGCTCGCCAGTTTTGTCCGCAGCGGCTGGGGAAATAATGCGCCACCAGTATCA
GCAGCAGATGTGGCTAAGGTCCGTGCCAGTCTTAATACCAGTCAGAAA

SEQ ID NO.80 - Orf 2035

ATGAAAAAATAACATTATTGTAATCAGCCGTGCTGGCGGGTCTGCTGGGCTGTACCGTG
GCACAGGCTGATGACAGTGGCGGACAACCTGGTCGCCCCGGGGAGAATATCTGGCGACTGCG
GGAGATTGTGTGGCCTGCCATACCGCCAGCGGTCCGGCTTTTACCGGAGGGTTGAAAATG
ACCACTCCGGTCGGCGCTATCTATTCGACCAATATTACCCCGGATAAACAGACCGGAATT
GGTGATTACACCTATGATGACTTTGCCCGCGCATTACGCCAGGGTATTGCCCGCGATGGC
CGTCATCTGTATCCGGCAATGCCTTATACGGAATATGCGAAGGTCAATGATGACGATATG
CATGCCCTGTATGCATACTTTATGCATGGTGTACCGCGGTACATCAGCCAAATAAACCG
TCAGATATTCCCTGGCCGTTAAATATGCGCTGGCCACTGGCGGTCTGGAATAAGTTGTTC
CTCGACAATACTCCGTTCAAAAACGACCCGGCACAAAGTGCTGAGTGGAACCGTGGGGCT
TACCTGGTCCAGGGGCTTGAGCACTGTGGTGCCTGTCATACACCGCGTGGTATTGCATTT
CAGGAAAAGGCTTCAGATGAGAAGGGAGCTGACTTCTTAACCGGTGGAACACTGGAAGGC
TGGCATGCGCCGGATCTGACCGGAAATGTAAAATCCGGATTAGGGCGCTGGAGCACCGGG
GATTTGCAGACGTTCTGAAAACCGGGCGCAATGACCAGAGCGCGGCATTTGGTTCGATG
AGTGAAGCCATCGGGCACAGTACCCAGCACCTGACCGATGCGGATTTACATGCTATGGCG

WO 2005/012547

PCT/US2004/024204

FIGURE 11C

GTCTATATTAAATCGCTAAAATCTTCTGATCCAGAGGCACAGCCTCCGGCGACCACCGAC
AGCACTACGGCGGCGTTAATCAGAGGAGATCTGAGCCAGACCGGTGCGGAAGAATATATG
GATAACTGTGCAGCCTGCCACCGTCTGGATGGCAAAGGGTATGCCAAAACCTTCCCGACA
CTGGCCGGTAACCCGGTATTACTGAGTGATGATCCTTCCTCACTGATTAGCATTGTCCTG
ACAGGCGGAAAGATGCCGGTGACTCAGCAATCGGTGACCGGACTGACCATGCCTGATTTC
GGCTGGCGGCTCAGTGACCAGCAGGTGCTGATGTGGTCAGCTTTATCCGCAGCAGTTGG
GGTAATAATGCCGGTAAAGTAGAGGCTAAGCAGGTAGCAGACATTGCAAGCTAATGCCG
GTACCGAATCAGGCAGATAATCCGCAGGTAAAGGCCGAAAAGCCGGATCCCGCTAAGAAA

SEQ ID NO.81 - Orf 1840

ATGAAAGCTATTAAAGGAATCATCGTTGTGATACTGGTGTGGTCATTATCCTTCTGGCC
TACGCTCTGTGGCCGACCAAAACAGCATCGCTTTCGCCGTTACCTGCGGATAACTCCCCT
CAGTTGGCCTCACTGGTCAGCCAGGGTCAGTATCTTGCGACTGCCGGTGACTGTGCGGCC
TGCCATACTCAGCCGGGCGGTAAACCGTTGGCCGGGGGACTGCCGATTCGCAGCCCGATT
GGGGTTATTTACACCACTAATATCACTCCGGATAAACAGACGGGGATAGGTAATTACTCT
CTGGATGATTTTGAACGCGCAGTACGCCACGGCATTTTGCCAAATGGCGACACCTTGAT
CCGGCCATGCCTTATCCGTCCTATGCAAAAATCAGCGATGATGATGTACGGGCCTTGAT
GCCTGGTTTATGCATGGGGTCCAACCGGTGAGCCAGCAGAACCGTGCCAGCGACATCCCG
TGGCCACTCTCGATGCGCCTGCCACTGGCGGTGTGGCGCAAGATGTTTGCGCCGGATCCG
GCTAACACCGGTTTTACAGCAGATAAATACCAGAGCGCCAGCCTGGCTCGCGGCGCTTAT
CTGGTTCAGGGCCTTGGCCATTGCGGTACCTGCCATACTCCTCGCGCCGGCACCTTGAG
GAAAAAGCGCTGGATGATTCCGGACAGCAGTATCTTGCGGGTGGTCAGGTGATTGACGGC
TGGCTGGCGGTAAATCTGCGCGGTGATAAAGCCGACGGTCTGGGTAACTGGACAGAACAG
GATATTATCGACACCCTGCGCACCGGGCATAACGTCAGCCATACTGTGGTGGGGCAGCCA
ATGGCAGAGGTTGTGGCTAAAAGCACCAGTCATATGAGTGATGCCGATTTGGCGGCCATT
GCCGCATATATCAAATCACTGCCTGCAGGCCAGGGTTCAAAGCATCGTACACGGAATCA
TCACAAACAGCAGACATGCTGGCCCGTGGTGAAAACCTACCCCGGGTGCGCAGTTGTAT
GTAGATAACTGTTCTGCCTGTCATCAGACCAGCGGTAAAGGTGTTTCAGCATATCTTCCT
GCGATGGCCGATAACCCGACAATACTGGCCGACAATCCGGTGTCGGTGATTGATCTGATT
CTTGACGGTAGCCGCTGCCAGCCACACCTCAGTCCCCGTGAGCACTGGCAATGCCTGGC
TTCGGCTGGCGTTTGTGTCAGATAAACAGGTGCGCGATTTAAGTAACTTTATTGCAACAGT

FIGURE 11D

TGGGGTAATAAAGCGACAGAAGTGACCGAACAGCAGGTGAAACAGGTCCGGGCAGATTAT
CCGCCGAAAGGCGAGAATAAGGATCCG

SEQ ID NO.82 - Orf 3651

ATGAAAAAAGCATATTAGCGCTGGTGTGTTGGTTCACTGGCTTTTTCTGCCATGGCCGAG
GATAACAGTGGTCAGGATTTAGTGAAGCGGGGCGAGTATCTGGCGCGGGCAGGTGACTGT
GTTGCCTGCCATAACCAGCGAAGGCGGTACGCCCTTTTGCCGGTGGATTGCCGATGGCAACA
CCTATCGGAAAGATCTATTCGACCAATATTACTCCCGATAAGACTTACGGCATCGGAGAT
TATACTTATGACGATTTCCAGAAAGCGGTACGCCACGGGGTAGCGAAAAATGGTGAGACA
TTGTATCCGGCAATGCCTTATCCGTCGTACGCAGTGGTCAGTGATGACGACATGCACGCT
CTCTACGCCTATTTTATGCAGGGTGTGAAACCTGTCAGCCAGCCTAATCACGCCACTGAT
ATTCCATGGCCTCTGTCAATGCGTTGGCCGCTGGCTATCTGGCGGGGAATGTTTGCTCCG
GCAGTAAAACCTGCCACAGCACAGCCAGGAGAAGATCCGGTGCTGGCGCGTGGACGCTAT
CTGGTTGAAGGGTTAGGCCACTGTGGTGCTTGCCATACTCCTCGTAGCATCACCATGCAG
GAGAAGGCGCTTAACAATAGCGAAGGTACCGATTATCTGTCTGGCAGCAGTGCTCCGATT
GATGGCTGGACAGCCATTAATCTGCGCGGCGACGATCGTGATGGTCTGGGCGGCTGGTTCG
ACCAGCGATATTGCACAATTCCTGCGTTATGGACGAAATGATCGGACCGCCGTATTTGGT
GGCATGACCGATGTGGTACAGCATAGCCTGCAATACCTGAGTGATGACGATATTAACGCG
ATAGCACGTTACCTTAAGTCTCTGTCAACACGGGATAGCCATCAGCCGGTATTTAAGGCG
GATGATTCTGTCTCGCAGGCATTATGGAAAGGTAATGATCAGCGAACCAGGGGCTGCTGAG
TATGTTGACAGTTGTGCAGCCTGCCATAAGACCGATGGTAGCGGGTACACCCGCTTCTTC
CCGGCGCTGAAAGGCAACCCGGTGGTACTGGCAGAAGATCCAACCTCACTTATCCATATT
GTTCTGACAGGGGATACGTTACCCGGAGTTCAGGGCGCGCCATCGGCGATCACTATGCCG
GCATTTGGCTGGCGGCTTAATGATCAGCAGGTAGCAAATGTCGTGAACTTTATCCGTAGC
AGTTGGGGAAATACCAGCACTGCGGCGGTATCGGCAGATCAGGTGGCTAAGTTGAGAAAA
TCAGCCGATGTGCAGGGAAAAATGGGTGATGCATCAGTAGAGAAATTACCTAAACAGCCT

SEQ ID NO.83 - Orf 3053

ATGGCTAAAAAACACGACGCGTTATCTCCGTGGTCGCTGCGGTAGTCATTGCCGGTGCA
CTCGGTTATACCGCCTATGAACAGTACGGTATTCATAAAAATCTATCCACAAACCGTCAGC

FIGURE 11E

CTGGAGACGGGCCCCGGCACTGCAGGACCAGATTAAACGGGGCGAATATATTGCCCGTCTT
TCTGACTGTACGGCCTGTCATACCGCTGAAGGTGGCCAGCCATTTGCCGGGGGCTATGCT
CTGCAAACGCCGTTTCGGGAAAATTCTGTCATCAAACATCACCTCTGACCGGGAAACCGGT
ATTGGTGGATGGACTCAGGAACAGTTTGATAAAGCTGTTTCGTCATGGTGTGGTTCTCAC
GGCTATCTGTATGCAGCCATGCCTTATCCGGCGTACTCACGGCTGACCGATGCAGACCTG
ACCGATTTATGGGCCTATATCCGTAACCTGCCAGCGGTTAACCATAAAGTGGTAGAAAAC
CAGCTGCCATTCCCGTTCAATCAGCGCTGGACGCTGGCGGGCTGGAATATGCTGTTCTTT
AAAGATGCGGCATTTACCCCTAATCCACAGGCCAGCGAACAGGTTAACCGTGGCCAGTAT
CTGGTCGATGGACCAGGGCACTGTGCTTCCTGCCATACTGCCAAAAATATGCTGGGCGGT
GACAGCTCTGCTTATCTGCAGGGCGGAGCATTGCAGGGTTGGTATGCACCGGACCTGACG
CCGGATCCTCATTCCGGTTTAGGCAACTGGAGTAATGCCGATATTGTCAGCTACCTGCGT
TCCGGCAGTAACCGTATCACCGCCTCCTCAGGCCCGATGACCGAAGCGGTAGAGAATTCA
ACGCAGTATATGAATGATAATGATTTGAACGCGATTGCGGCTTATCTGAAATCTATTCCT
GCCTCTCACCCACAGGTTCCGACAGCTCTGACGGCTGATGACCAGCAAATGGTCTCCGGC
AAGAAAGTGTTTGAATCTCAGTGCAGTGCCTGTACGTTTCTGATGGTGCAGGGATTTCGC
AACATGATCCCGGCACTGGCAGGCAATCCACAGGTAAATTCTGCAGATCCGTCCAGCCTG
CTAAACGTGGTACTGAATGGCAGCGAGGGACCGTTTACCCATGCGAATCCGACGGCTGCA
GGTATGCCATCGTTCGGCTGGAACTGTCTGACGCTAATATTGCAGAGGCCCTGACCTAT
ATCCGTAACAGCTGGGGCAACGCCGCTCCGGCCGTACGGCTGACCAGGTGAGCGCAGCC
CGTAAAGCAACCGGAGCAAAAAGCTGGCTGGGAGATTCCATCGCCTCTCAGGACAGCGGT
AAA

SEQ ID NO.84 - Orf 2632

ATGAAAAAGACAACAATAGCCATTGCTGTTGCGGGGATCGTCGTGGTTCGGCGCACTCGCT
GCATTATGGATGAACGGCAGCACGCGGGCAGACGATGTAGCCGGAGACCAGGTGCAGACC
AGCCAGCCGGTGTCTGCAGAAGACAGTGCGGCGGTGAAACGGGGTGAATACATCGCAGTA
GCCGGTGACTGTGTGCCTGCCATACTGCCCCGGGAAGTAAAACCCCGTTTCAGTGGCGGA
TATGGGATTGATACGCCGTTTGGGACCATTTATGCCAGTAATATCACGCCTGATAACCAG
ACCGGAATTGGCCAGTGGACCGAACGTGATTTCTACCGGGCAGTTCGTACGGTATTGGC
CGGCAGGGCGAAAATCTCTACCCGGCGATGCCCTATAATGCCTACGTGAAGGTCAGTGAT
CAGGATATGCATGATCTGTGGATGTATATGCGTACGGTTAAACCGGTGAATCAGCAGCCG
CCGGAACTCACCTGCCGTTCCCTTATAATATCCGACTGGCAATGCGTGGCTGGAATCTG

FIGURE 11F

CTGTTTTTTTAAAAATAGTGGATTTGATGCGAATAGCAGTCAGTCAGCAGAGTGGAACCGC
GGGGCTTATCTGGTTCAGGGGCTTGAGCACTGTGCTGCCTGCCACACGCCGAAAAATATG
CTGGGTGGAGATACGTCAGCGTATCTGCAAGGTAGCAGCCTCGGACAGTGGCATGCACCG
GAAATTACCGGCAACACCTATAACCGGTATTGGCCAGTGGAGTGAGCAACAGGTGGTTGAT
TACCTGAAAAGCGGCAGCAATCAGGTTGCAGTAGCTTCCGGACCGATGGCCGAAGCGGTG
ACCAATTGACTCAACATCTGACGGATGCTGATCTGCGGGCCATTGCAGTCTATCTGAAA
TCCCAGCCGGGTTCCGGCAAACCAAAGCCAGCGGCTTTGGCAGCAACCAGCCCGTTGATG
CAACAGGGAGCGAATGTTTATCAGGCAAACCTGTAGTGCCTGTCATAACAGCGATGGCCGG
GGAATACCTCAGCTGGCTGCCGGGTTGCGGGATAACCCGGGAATAATGGCGGCCGACAGC
TCATCGGTCATTACTACTATTCTTGAAGGCGGGCGCGGCGCGGTGACCCTGAACAATCCG
ACCAGTGGTGCTATGCCCTCGTTCGCATGGAACTGTCTGATCAGCAAATTGCCGCGGTA
TCTTCCTATATTCGCAACAGCTGGCAAATGCGGCCCCGGCAGTGACTTCACAACAGGTG
GCTGCAATGCGTAAGCAACTGAACTGACCCCTCAGTTACCAGATAACGGAGAGCCGGCA
CAT

SEQ ID NO.85 - Orf 1750

ATGACGATTAAAAAATATATTGCTTCAGTTGTCGGAGTGGCTGTAGTGGCCGGACTCGGA
TTCAGTGGCTGGAAATGCTGGCATAACGCTCATCAGGATCACAGTTTTGTAGCCCCTGCA
TCTGCCGGGGATACCGGCAGCACGGCGATTGCACGTGGTAAATATCTGGCCACGGCCGGG
GATTGTGTTGCCTGTCACACTGCGCCTGGCGGTAAAGCCTTATGCGGGTGGACTGGGACTC
AATACGCCGTTTGGTACCATCTATGCGACCAATATCACCCCGGATAAAGAGACCGGAATT
GGCGGCTGGACTGATCAGCAGTTTATGAACGCTGTTTCGTAACGGTAAAGGTGCTAACGGT
GAGAATCTCTATCCGGCAATGCCTTATAATGTTTATGCTCAGGTCAGCGATCAGGATCTG
AAAGATATTAAAGCCTATCTGGACAGCGTACCGGCAGTACATTACACCGGACCAAAAACA
GATTTACCATTTCCGTATAATATCCGGCTGATGATGATGGGCTGGAACCTGCTGTTCCCTG
AATACCGCAGCGTTCAAAGCTGACCCTGCACAATCTGCCCAGTGAACCGTGGCGCTTAC
TTAGTTGAAGGGCTCGGGCATTGTACTTCCTGTCACACACCAAAGAATATGCTGGGTGCG
GATAAAATGGGGGTTCATCTGCAAGGTGGTGAGCTGGAAGGGTGGCTGGCTCCGGAAATT
ACCGGCAACACGCGTCAGGGAATCGGTGGCTGGAGTGATGATGAATTAGTGCATTACCTG
AAAACCGGAGCAAACGATAAAACGGTTGCTGCAGGTCCAATGGCTGAAGCGGTTTCATAAT
TCACTGCAACATTTGAACGATCAGGACTTAACGGCAATGGCAACTTACCTGAAAAGCTTG

FIGURE 11G

CCAGGCAGTGAAGACAAATCAGTTGCTCTGAGTGGAATGGATGATGTGATGGCCCGCGGA
CAGAGTATTTATCAGGCGAACTGCTCTGCCTGCCATCAGTCGGATGGTGCTGGTGTCCGG
GATATGGTGCCGGCGTTAAGGGGTAACAACGGGCTACAGGCATTTGAACCGACCAACGTG
TTGCATGTACTGATGATTGGGGCACAGGGCGCAGCCACCGCCAGCAATCCTACCAGCGCG
GCGATGCCAGAGTTTGGCTGGAACTGACTGACCAGCAAATGGCAGATGTCAGTACTTAT
GTCCGTAACAGTTGGGGAAATAAGGCGCCAGCTGTCACCGCATCACAGGCGGCGGCTGCA
CGGAAACTGCTGTCAGGTTACCGGCATTGCATAATCCAGCGGCAAAC

SEQ ID NO.86 - Orf 1953

ATGATGAAAAAGTTAATGCTGACTGCCGGCAGTTTACTGTTGCTGACAGCCGGTTACGCA
CATGCTGACAGCGGTGGCGATTCTTGGGACCTGGTAAGTAAAGGCCGTTACATTGCGCAA
CTGGGTGACTGTACCGCCTGTCATACAGAGCCGGGACATCCTCTGTTTTCCGGCGGGGTC
GCGATTGAGACACCTTTTGGAAACTGGTAGGCGCGAATATTACCCCTGATCCGGAAACC
GGCATTGGTAAATGGACCTTTGAAGACTTCCAGAACGCGATGCGTAAAGGCCACAGCCGT
GACGGTCAGTTGCTTTACGGTGCCATGCCTTTCACGGCCTACACCAAAGTAACTACCGAC
GATAACCGGGCACTGTGGTCTTATCTGCAAACCTGTTGAGCCGGTAAACCGGGTGGTTAAC
ACCAACCAGTTGCCATTCCCGTTCAATATCCGTACTTCACTGCATGTCTGGGATATGCTG
AATTTTACCGAAGGTGAATATAAGCCGGACCTAAACAATCAGCCGAATGGAACCGGGGC
GCTTATCTGGTTCAGGGTCTGGGGCATTGCAGTACCTGCCATACACCTAAAAATATGCTG
GGTGGTGATAAAGACAGCAAGTTCCTGCAAGGCGGCTCACTGGGTGTCTGGTTTGCTCCG
GATATTACCGCCAATACCCACAGCGGTATTGGTCAGTGGACCCAGCAGGAAATTGTCGAA
TACCTGAAAACCGGTGCTAATAAATACGATATCGCTTCAGGTCCGATGGCTGAAGCTGTA
GAGCATTGACTCAGTACTGGAAAGATGAAGACCTGAATGCGGCTGCGGTGTACCTGAAA
TCGCTGAAAACGATAGTAGCCAACCACAACCTCTGGCGGCGGATAATGGCCAGATGGTG
AATGGTAAAGCGATTTACGCGGACCGTTGCAGTGCCTGCCATGTGTGTCACAAGGTCAGGGA
GTCTCACATCTGTTCCCGCAACTGGCTAATGCACCACTGGTTAATGCAGTCGACCCTGCA
TCACTGATTCATGTAGTGCTGGCGGGCAGTCGCGCCGGAGGGACCGCTGCGGCTCCAACG
GCTCCTGCTATGCCTGCATTTGGCTGGAACATGACGGATCAGAACGTCGCCGATGTGCTG
ACCTATATTCGTAACAGTTGGGGGAATGCAGCACCGTCTGTCACCGCCAGCGATGTGAAG
AATATGCGCAGTACTTTAGAGAAG

FIGURE 11H

SEQ ID NO.87 - Orf 3687

ATGCAGAAATTAAGGGTGTTTACCCCGTTGGCTATTATGCTGGCTGGGTTTTGTGGCTCTGTTTA
CGCTGATAACAGTCCTGCCTCGTCAGACAGCACATCGCTTCCC CGGAGAATATCTGGCCAGAG
CCGGCGACTGTGTTGCCTGCCATACAGCAGAAGGAGGCAAACCTTTTGCCGGTGGATTGAAAATG
ACGACGCCGGTAGGGGCCATCTATTCAACAAATATCACTCCGGATAAAGATAACGGGATCGGTAA
TTACAGTTACGATGACTTTGTCAAAGCTGTACGCCAGGGAGTCAGTAAATCCGGATCAACCCTCT
ATCCGGCGATGCCGTATGCTTCTTTTACCAGGATATCAGACCAGGATATGCATGATCTGTATAAC
TATTTTATGCAGCAGGTAAACCGGTCAGCCAGCAGAATAAAGCCTCTGATATCCCCTGGCCTTT
GAGTATGCGCTGGCCACTGGCATTCTGGCGCTGGACATTTACCGATGATAAGCGTTTTTCAGCCTG
TCGAAGGTAAATCGGCAGAATGGCAACGGGGGGCGTATCTGGTCGAGGGGCTGGAACATTGCGGA
GCCTGCCACACTCCGAGAGGTATAGCATTCCAGGAGAAAGCACTTGATCAAAGTGATCCGGTTTA
TCTGACGGGTAAACACACTGGAAGGATGGTATGCACCGGATCTCACCGGCAGCAATCTGATGGTC
TGGGTCGCTGGTCACAACAGGACATTGTCTAGTTTCCTGAAAAATGGTGTAACGGCACAAGCTCT
GCCTTTGGTTCCATGTCAGAGGTGGTTCATGACAGTACCAGCTATCTTACCGACAGTGATCTTCA
GGCAATTGCAGTCTATCTGAAATCGCTGCCTGCGGCACACCAGACGCAGGCGCCAGCCAGTAATA
ATGCTACCGCTCAGGCACTTTTTTAAAGGCGATGTTTCTGCTACGGGTGCACAGGTTTATCTGGAT
AACTGTTCTGCCTGTCATCGCTCGGATGGTAAAGGGTATGATAAACGTTCCCGTCACTGGCAGG
CAATCCGCAGTACTGAACAGTGACCCTTCATCAGTGATTCATATTATCTTGCAGGGGGGACAAC
GCGCAGTGACACCAGATATGCCGACCGGATTAACCATGCCGGACTTTGGTTGGCGGTATCGGAT
CAGCAGGTCGCGGATGTCGCCACCTTTATCCGT CAGGGATGGGGGAATAATGCTGCTGCGGT CAC
AGCCAGTCAGGTTGCCGATATCCGCAAGCTAATCCCGAAACCCGCTTCTCAGGCTGCTAAGTAGC

FIGURE 12A

SEQ ID NO.88 - Orf 3820

ATGAGCAGGAGTGTAAGAGTGAGACCAACGAGTCTGGCTTTAATTATCGGGCTCTCGGTG
TTCTCGGGGAAAGCTGTGCAGGCTGCAGACACCCCTTCAGCATCGACGATAATTGAGCAA
GGAAAATATTTATCAGTGGCCGCGGACTGTGGAGCCTGTCATAACTCCCCGACAAGCGGA
GCTGCTATGGCGGGGGGCTATGCGATTGCCTCACCAATGGGCAATATTATCGCCAGCAAC
ATTACCCCGTCAGTGACAGCCGGCATTGGTAATTATACCGAACAACAATTTGCCCGGGCG
GTCAGAGAGGGAGTTAACGCACAGGGCGACCATCTTTATCCGGCAATGCCTTATACCTCG
TACAGTAAAATGACTGACAGTGATATTCATGCGCTGTATCAGTATTTTATGCACGGGGTT
CAGCCGGTTGATACTCCGGCTCCGGCCACAAAGCTTCCGTTTCCGTTCTCAATTCGTAGC
AGTATGGCGTTGTGGAATATGCTGTTTGCCAGCCAGCAGCGTTTCACTCCGGATAGCCAG
AAATCAGCTCAGCTAAACCGGGGTGATTACCTGGTCAATGTGCTGGAGCACTGTGATGCC
TGCCATACTCCTCGTAATTTCTGATGGGTGAGAAAATGACCTGGCTTTATCCGGCGGG
CAGGTGGGTAGCTGGTATGCTCCTAATATCACTTCTGATAAACTGCCGGTATTGGTAGC
TGGAGCGATGACCAGCTGTTTCAGTACCTGAAAACAGGTCATGTTGCCGGTAAAGCTCAG
GCCGCGGGCCCTATGGCAGAAGCCATTGAGAACAGCCTGCAACACCTTAGCGATGATGAT
TTGCATGCCATTGTTGCCTGGCTGAAACAGGTTCCCGCCTCGGGCGCCACAGCTACGGAA
TCACGTTTTACTCAGGGTGCGCCTTCAGACAGTGAGGCCGCCATGCGCGCGACCGATCAT
CCGGATGCCGGCTGGGTGCTGTTTCAGTAACAGCTGTGCTAACTGCCACCAGGCCAACGGT
GAAGGCAGCCAGTTTTATCCTTCGTTATTCCACAACAGTGCAACCGGTGCCGCACAGCCG
GACAACCTGATTGCGACCATTCTGTTTGGTGTCCGCCGTCACGCCGACGGCCAGTATGTT
GCGATGCCAGCATTCCGACCTGCAGCTTCGTTTGTGACCGGCTCAATGATCAGCAAGTT
GCAGATGTGGCTAACTATGTCTTAAAAAATTACGGAAATGCCTCACTGACTGTCACTGCC
GATCAGGTGAAAACAGTTCGTGAAGGCGGGCCTGTACCGGCCATTGCTTATCTGTCAAAT
CCGGCAGTGTTAGCTATCGGTGCATTGATTGTGCTGGTGATTCTGGGCCTGATCGTGACT
GCAGTTCGCAGAAGGGGGAAAAAA

SEQ ID NO.89 - Orf 1956

GTGAAACAACAACAAGCTAAACGCGCATAAAGCCGCAGGTTTCCGCCGAAAGCTACTG
AGTCTTTGTCTGGGACTGAGTGCATTAAGCGCGGTTCTGTGATGGCAGCTGAGCAGGTG
CCGGTCAGTCAGCCTTCTGTGGATAACAGTGCTGATGCATTGTTGAAACAGGGGCATTAT
CTGGCCATTGCTGCTGACTGTGCCGCTGTCATACCGATCCTCAAACCAAAAAGACTTTC

FIGURE 12B

GCCGGTGGTTACGCCATTCACTCTCCAATGGGGGTGATTTATTCCACCAACATCACTCCT
TCACGGCAGTATGGTATCGGTTCTACAGCGAAGCTCAGTTTGAACAGGCAGTTCGCCAT
GGTATTCGCGGTGACGGGAGCCACCTGTATCCGGCCATGCCATACTTCCTATTCTGGGT
CTGACCGATCAGGATATTCATGCGTTGTATTACTATTTCACTCACGGCGTACAACCGGTA
GAACAGGCTAATCGTCCGACAGAACTCAGCTTTCGGTTTAATATCCGCGAAGCTATGTGG
GGCTGGAATTTACTGTTCTGAAACAAAACCTTTCGGTGACGACCCCTCCCAAAGCCCG
CAATGGAACCGTGGTAAGTATCTGGTCGCGAACCTTGAACACTGTGGAGAGTGTACACC
CCACGTAATACATTGATGGGCAGCGAAACCGGGTCGGCACAGTATAGCGGTGCAGCCCTG
GGAAGCTGGTTTGCACCAAATCTGACTTCTGACCAGCAGAGCGGTCTGGGCAGCTGGCAA
CGTGATCAGCTCATCACCTATCTGAAAACAGGTCATGTTGCCGGTAAAGCTCAGGCTGCC
GGACCGATGGCGGAAGCTGTCACCAATAGCCTGCAATATCTGAGTGATGATGATATCGGG
GCCATTGTGACGTATTTGCAAAGTCTGCCACCGGTCAGTGAACCAGACCAGGCGAAAGCT
ACCGGTGATTTCCGGCAGTTCCGCTGGTAACAGTTCAGATTCTGAGGTTCGTGGTACTCAG
CCTATGGGATCTGTACTGCCGGACGACATTACCGGTAAAGCCCTTTACGACACAACCTGT
GCCAGCTGCCATCAGTCTTCCGGAGCGGGTACTACGGATAATTTCTATCCATCACTGTTC
CACAACACTGCCACCGGCGGAAATACCCCGAACAATCTGGTCTCCGCGATTCTGTTTGGT
GTGCAACGTGAAGTGAACGGCAAGCAGGTACTGATGCCAGCCTTTGGTCCGGGATCAGAC
GTGCAATCGCTGAATGATGAGCAGGTAGCAAACTTAGTAATTACATCTTCAAACAATTC
GGCAATCCACAACCTGTCGGTGACTGCTGATCAGGTAAAAACACTACGTGAAGGCGGCCCA
CAGCCGTTCTTCGCCAAATACGCTGCATCAGGATCGGCAGTGGGTGGTGTGATCCTGTTA
CTGATTATTGTGCTGATTATTGTCCGCATTTTCGCGCAAACGTCGC

SEQ ID NO.90 - Orf 2446

ATGAAATGCGCTTATTTATCTTTACTGATAAGTACGCTGCTGTATGCCGGTTTTTCGCCC
GCCACTCAGGCTGAAACACCCGCGACAGCAGAACTCTGCTTGCACAAGGAAAATATTTA
TCTGTGGCTGCAGACTGTAGCGCCTGCCATGACAGTCCCGATCACCATGTCATGGCTGGC
GGAAACAGTATAAATTCCCCACTGGGCAAGATTGTTGCCAGTAACATTACCCCGTCAGTG
CACTATGGCATTGGCAGTTATACCGAGCAGCAATTTTCGGATGCGGTTCGAAAAGGTATT
AATGCTCAGGGCGAAAACCTCTACCCTGCAATGCCTTACACCTCTTACAGCCAGCTCACC
GACAGCGATATCCACGCGTTGTATTACTATTTTATGCATGGTGTAAGTCTGTGATCGC
GCCGCTGGGGCCACACAGCTGCCCTTTCGGTTTAACCTGCGTATCAGCATGAAACTGTGG
AATGCTCTGTATGCGGACAACAAGCCGTTCCGTCCTTCATCCTCGCAAACCGATCAGGTT

WO 2005/012547

PCT/US2004/024204

FIGURE 12C

AACCGCGGCAACTATCTGATTTACGGGTTGGCTCATTTGTGATACCTGCCATACACCGCGA
AATGCCCTGATGGCCGAAAAATCTGACCAGTCTCTCTCCGGGGGATCACTCGGCCAATGG
TATGCACCCAATATCACCTCAGATAAGTCCTCTGGTATCGGTAACTGGAGCGATCAGCAA
CTGTACCAGTATCTGAAGACCGGACATGCTGTCTGGCAAAGCCCAGGCTGCCGGGCCAATG
GCAGAAGCTATTGAACACAGCCTGCAATACCTGTCTGATGATGACCTGCATGCGATTGTA
GCCAGCCTTCGTCTAACCAGACCGGTGAATACTGCATCCGCAGATCGCGGGATGCAGGGT
AAAGCAATATCTGATGAAAACAGTATCCGTGGTACCAAAGTTGCCAGCGGAGAGCCGGTT
TCCGGCCCAATGTCAGGGGCAATTCTGTACTCAGGTAAGTGTGCTGCCTGCCATACGCCT
TCCGGTGCCGGATCCTACAGCCAGAATTATCCGTCACTGGTACACAACACCACTGTAGGG
AGCACCGATCCCATAATCTGATTGCCACTCTGTTATTTGGCGTTCACCGTACCGTTGAT
CAACAAAGCATCACAATGCCCGCCTTTGGGCCACAAGGCTATACCGACCGTCTTAGCTTT
GCAGAGATTGCCACACTGGCGACCTATGTCCGGCAGACTTATGGCGCCGGAGGTGAAGCT
GTCAGCGAACAACAGGTAGAGCAGGTTTACCAGGGTGGGCCTAAGCCGCTGATTGGCTGG
CTGGCCGACGGAAGAATACAGGCTTTAATCGTTGTAGTGCTGTTGCTGTTGGCTGGCCTG
ATTATCACCGTAGTGCGCAAAGGGAGAAAAGCA

SEQ ID NO.91 - Orf 3674

ATGAAAAACATGCCATAAAGTTTTCCCTGTCGCTAATGTTTGCCGGAAGCATGTTGTGG
GCTGGTTCCGCTGCAGCCGCAACTGGCGATGCCGCGGCGGCCATTAGCCGTGGTGAATAT
CTGGCGACGGCTTCGGACTGTGCTGCCTGTCATACTGATAAAGGCGGTCTGCCGTTTGCC
GGTGGATTAAAAATTGAATCCCCGGTCGGGACGATTATTGCCAGTAATATTACTCCCTCC
TTAACCGCTGGGATTGGTCATTATACTGAACAGCAGTTTGCGGATGCGGTGCGTAAAGGG
ATTCGTGCCGATGGTGCTAACCTTTACCCGGCCATGCCTTATACAGCGTACTCGGTGATG
ACAGATCAGGATATTCATGATCTGTACCAGTACTTTATGCAGGGCGTAAAACCGGTGGAT
CATCCGGCCGCGGAGACTGAGTTACCGTTCCCGATGAACATCCGCATGATGATGAAGGCC
TGGAATTTACTGTTCCCTCAACGATAAACCATTTCAGCCCGGATGCATCACAAAGCGCTGCG
TGGAACCGGGGTAAATATCTGGTAACTGGTGCAGCGCACTGTAGTACCTGTCATACACCC
AGAGGGCCACTGATGGAAGAGGAAAGCAGTCAGTTCCTGAGCGGTGGTCAGGTCGGGGCA
TGGTACGCGCCAAATATTACTTCGGATCCACAGTCAGGGATTGGCCGCTGGAGCCAGGCT
GATATTGTACAGTATCTGCGCACTGGTAATCTGCCAGGTAAAGCCCAGGCTGCCGGTAGT
ATGGGAGAAGCGGTAGAGCATAGCTTCCAGCATCTGACAGATGATGACCTGAATGCGATT
GCTACCTATATTCGTACCGTGAAGCCAGTGGCAACGCCCCGAAAACGCCGGTTCAAGATTT

WO 2005/012547

PCT/US2004/024204

FIGURE 12D

ATGCAGGGTGACAGCCACGATGCTACCGGAAAAATTCGTGGTCTGAGTCAGCAACAGGTC
ACTGATGCTAAACAGCAAGGGCTGGCGTTGTTCCAGGGCAATTGTGCCTCCTGTCATGAA
GCTGGCGGGCAGGGAAGCAGGGATAGTTATTATCCGAGCCTGTTCCACAATTCAGTGACC
GGCGCAGAGAACAGTAATAACCTGATTGCTACCATTCTGAATGGCGTTAACCGTACCACC
CGGGACGGTCAGGTATTTATGCCAGGTTTTGGTCATCATCCAAATGATATCAATAATCTG
ACTGACGAGCAGATAGCGTCGCTGGCAAACCTATGTGCTCACAACCTATGGTAAACCGTCG
AAACCGGTGACTGCGGCGATGGTTGCCACCGTACGCCAGGGAGGGCCGGGTTCCAGTCTG
GTGCTTCTCGCCCGGTTTGGGAATAGCTGCCGGAGTGGTAGTTGTTTTGATTCTGCTGGGA
TTCTGGGTGGTTCGCCGCAAAAAAACGTCAGGGATCCGTCG

SEQ ID NO.92 - Orf 1221

ATGAAAAAACTGCTTTCCTGTGTATCGCGGGTGCCCTGGCCGGGATCATGCTGAATAGT
GCCGCCATGGCTGAAGACAGCAATGCTCAGAGCCTGATCGCAAAGGACAGTATCTTTCG
GTAGCCGGTGACTGTGCTGCCTGTCATACCACCAGCGGAGGAAAACCTTTTGCCGGCGGT
CTGGCTATTGCCACACCGATTGGCAAGATATTCTCCACCAATATCACTCCGTCAAAAACC
TCCGGCATTGGCGATTACTCGCTGCAGGAGTTTGAAAAAGCGGTTTCGTCAGGGAGTAAGA
AAAGACGGCGCCAATCTCTATCCGGCGATGCCTTACACTTCTTATGCCAAAATTTCCGAT
GAGGATATGCAGGCTCTGTATGCTTATTTTCATGCATGGCGTTGCGCCTGTGGATGAGAAA
GGCCCGCAAACCTGCCCTGCCATTCCCGTTCAATATTCGTCTGTGATGGCGGGCTGGAAC
CTGATTTTTGCCGGAGACAAACCATTACGCCAGACAGTAACCAGTCAGCAGAATGGAAC
CGCGGCGCTTATCTGGTTCAGGGTCTGGCTCACTGCTCCACCTGCCATACCCCGCGTAAT
GCTTTGATGGCTGAAGAGTCTGGCCAGGCGCTGGCTGGTGCTTCTCTCGGTACCTGGTTT
GCACCAAATATTACCCCGGATGCCCATGCAGGGATTGGCAAATGGTCAGCCAGTGATTTA
GCCACTTATCTGTCTACCGGTCGTTACCGAACGGTTCTCAGGCTGGCGGCCCCGATGCTG
GAAGCTATCGATAAAAGCTTTAGTAAACTTTCTCAGTCAGATATTAATGCGATAGTGACC
TATGTGCGTTCCGTGAAGCCTCAGTCAGCCAATGCCGCTCCGGGCCAGGTACCTGCCAGT
GCCCCGGTAGTGTCCGATTTTGCGCTGATGAATGGAACAGCTTCTGATGGCGCGAAGCTG
TATGAAGCTCACTGCTCCACCTGCCACCAGGCCTCTGGTCAGGGCAGCAATGGCTTACCG
GCTTTATATGGTAACGCCGCGCTGCATCGTCCGGTAGCGGATAACGCAGTCATGGCTATC
CTGGACGGCCTGACTCCGACTCAGGGCCAGGCTATGCCGTCGTTTAAAACTGCCATGAAT
GATCAACAGATTGCCACCCTGACCAACTACCTGTTTAAACCTTTGGCGATGCCGGTGTT
CAGACCACCGCAGACAGAGTTAAGGTATTACGGGAAGGTGGAGCACCATCTCCGTTACTG

WO 2005/01254/

PCT/US2004/024204

FIGURE 12E

GCGATTGCCAAAGGCGGGATGATAGCTGCGGTGATCGTGGTATTACTGCTGATAGTGGGC
GGAGTGATGGTTAAATCGCGGCGTAAACGCCCGT

SEQ ID NO.93 - Orf 1984

ATGAAAAATACTCAGCTCTTCTGACTTTGTCAGCTGCATTCTGTTCTCCCCCTCGCT
CTGGCGGCCACCAGCAGCAATAGCGATTTAGTCAGTCGCGGTGAATATCTGGCGCGGGCC
GGTGACTGTACTGCATGCCACACTGCCGCAGGTGGTGCCGAATATGCCGGCGGGTATAAA
TTTAATATGCCTATGGGCACTATCGTAGCACCGAATATTACCTCTTCAGTGCAATACGGT
ATTGGTAACTGGTCTGAAGCCGATTTTGCCAAAGCAGTAAGGCAAGGGGTACGCCCTGAT
GGTTCTCATCTCTATCCGGCAATGCCATATACTCTTACGCCACAGTTACCGATGAAGAT
ATGCAGGCGTTATATGCTTTCTTCAAACCGTTCCGGCAGTAGATAAAGCCCCGGCAGAT
AAAAACGACCTGAAATTTCCGTTTAACTGCCAGGCCTGATGGGGATCTGGAATGCTTTG
TTTGCCAGTGATGCGCCATTTAAAGCCGATCCGGCATTAACTGCTGAGCAAAACCGGGGG
AAATATCTGGCCGAAGGGCTGGCTCACTGTTCAACCTGTCACAGTCCACGCAATCAGATG
ATGGCCGAGGATACTCATCAGTTGCTTGCAGGCAATCATGTGGATGGCTGGCTGGCACCA
AACATAACCTCTGATGCTGTCAGCGGCATCGGTGGCTGGAGCCAGCAGGAAGTACCGAA
TATCTGAAAACCGGCCATGTGGAAGGGAAAGCTCAGGCCGGTGGTCCTATGGCTGATGCC
ATCGAGCACAGTTTCAGTCACTTATCAGACAGTGATTTAGCCAGTATTGCCACATGGCTG
AAAACAGTACCAGCCATCCGCACTCCGGGCCAGACACAACCTTCATGGGCTGCCGCGCCA
GCCAGTAAGGTAGACTGGACAAGTTATCAAACCGGGGGCGGGAAGAATAATTCTCCGGCG
TACCGTGACTCGTCCACTACCGACGGAGCCGTACTGTTTGACAGCAGTTGTGCGGCCTGC
CATCAATCGAGTGGCCAGGGTTCAGACGACCATTACTTCCCGTCTCTGACCCATAACAGT
GCGGTTGGTGCAGCGGACCCGTCAAATCTGGTTATGGCGATTGTTGATGGTATTCACCGT
AAAACCCCGGAGGGTGAAGCGGTTATGCCAGCGTTCTCTTCAGAACTCAGGCCATTAC
TCATGGCTGAATAATGATCAGATTGCGGCAGTGACTAACTACGTACCGAAAAATTTGGT
CACGGAAATGCCGGTCTGACCGGTGCCGATGTTGAGAAAATCCGTAACGGCAACAGCAAT
GTTCCGTTCCCTGATTAAAAATGCAGGCGGCCTGACCATTGGCGGGATAGTGATTGTCGTT
ATTATCATTATTGCGTTACTGGCGGCACGCAGCCGTAAAAAACGTCGA

SEQ ID NO.94 - Orf 3400

ATGAAAGCTGTGATAATTCGATCGGCCATCGCGTTAGCTCTGATGCATGGCAGCCTGGCA

FIGURE 12F

CTGGCGGCAGATGATAATGCTGACTTAATTAAACGCGGAGAATATCTGGCAACTGCCAGT
GACTGTACCGCTTGTCTACACGGCCCCCTGGTGGTCCGGCATATGGCGGTGGTTACCCGGTA
GCGACTCCGTTCCGGTAAAATCTGGGGAAGTAATATCTCCTCAGATAAACAATTTGGTATT
GGTAGCTGGACCGACGATCAGTTTGTCTCGCGGCGGTCCGTGAGGGTGTCCGGTAAAAACGGC
GAGCAGCTGTATCCGGCGATGCCGTATGACGCATTTACCAAGATGAAACGCGACGATGTT
CTGGCCATCAAAGCCTACCTGATGTCATTACCGGCTGTGCATAAAGCCGCACCAGAAACC
TCTCTGCCGTTCCCGTTCAACCAGCGTTGGGGGATGCGTTTCTGGAAAATGTTTAACCTG
ACTGAAGGCGAACTGAAAAATGACCCTCAGCAGAGCCCACAATGGAACAATGGTCGTTAC
CTGGTAGAGGCACTGGCACACTGTACCACCTGCCATACACCGCGTAATCTGACCATGGGA
ATGGATACCAGCAAACCGCTATCCGGCGGGGATTTAGGCGACTGGATTGCCTTTAATATT
ACCCCGGGTAAATCAGGAATCCGGCGACTGGTCGAGTCAGGATATCGTGACCTACCTGAAA
ACCGGTTACCTGGCAGGCAAAGCCTCAGCATCAGGCCCGATGGCCGAAGCGATTGAGCAC
AGCCTGCAATATCTGCCGGATTCAGATTTGCAGGATATAGCCACTTACCTGAAATCGGTA
AAACCGGTGGATGATGAGAAGCAGAGCGTGCCACGGGACAGCCAGGGACAGCCGTCGGAT
GCCATTATCCGCCTGCGGGGTGCTGATGCAGCCACTCTACAGTCACAACCCGGCGCTGTG
GTATTTGAAGGCAACTGTTCTGACCTGTCACGGTGCTGAAGGTGCCGGATCAGGCCAGGGA
TTCCACGCCTATCCGTCACTGTTCCACCACTCCAGCACCGGAGCCATCGATCCTAAAAAC
GTGGTATCGGTGATACTGAACGGTGTTAACCGCCATATGCAACAGGGTGATATTTTTATG
CCATCCTTTGCTCCGCAGTTAAACGATCAGCAGGTGGCTGATGTGGCGAACTTTGTGATG
CAGAAATTTGGCAATCCGGCGGGCTGAAAAAGTAGACACCAGCCAGGTTTCAAAGCACGG
AAAAATGCCAGTCTGCCGTTACCGCCAACCTTTGCAGACGGTGCTAATCCG

SEQ ID NO.95 - Orf 3041

ATGATTCGTTTCATCTTTTAAGCGCTCACGGAATTTTCTGCEGTTGGCEGGGCTGTTGTTT
TGTGCTGCGGGCTATGCGCAAACCGGATCAGCTCAACCTGATCCGGTGGCAACACAACCC
ACTCCGACTCAGCCAGCGGCCGCTGCCGGTACGCAGGGAACGACCCTGATTCAGCAGGGT
GAGTATCTGGCGAAAGCCGCGGACTGTGAAGTTTGTCTACACGGCTACCGGCGGGCAGACC
TTCGCTGGTGGGCTTGGGTTTAAACCCCGTTTGGCACCATTTTTTTCATCGAACATCACC
CCGGATAAAACCCACGGGATTGGTCAGTGGAGTGAGAAGCAGTTCAGTGATGCGCTACGT
TATGGCATCCGTGCTGATGGCAAGAATCTGTACCCGGCAATGCCTTATACCTCTTATTCT
AAGCTGACGGATGCAGACATACATGCCATGTATGCCTTCTTTATGTCGCTGAAACCGGTC
GCCACCGATCCCCCGGAAAATAAGATGGGCTTCCCGTACAATCAGCGTATTGCCCTGAAA

FIGURE 12G

GGCTGGAATCTGATCAATTTCCATTACCAGCCGTTTAAGCAGGATCCTGATCAGTCGGCC
GAATGGAACCGGGGTCACTACCTGGCAACGGCGCTGGGACACTGTGAAGAGTGCCATACT
CCACGTAACTGGCGATGGGCCTGAGCGATAAGTCCTATGCCGGTGCGATGGTGGATGGC
TGGGAAGCATTTAACATCTCTTCGGATAATACTTCCGGGATTGGTCGTTGGAGCCACGCT
GATCTGATGCAGTACCTGAAAACCGGTTCTGTACCAGGTGTGGCTACCACGGGCGGAGGT
ATGGCTGATGTTATTTCTCACAGTCTGCGTTTCCTGAGCAATGATGACCTCAGTGCACTG
GCAACCTATATCAAGAGTGTTCCGCCGCAGAAACTGCGTCACAAAACCGCTCCGGATAC
GGCGACAATGTTTACAGAGTGATATTACTCAGGCGGTTTCGCGGCATGCCAATTGATGATTCC
GCACCTTCTGGTGCCGTATTATTTAACGGCAACTGTGCCAGCTGTCACGGGACCAAAGGT
CAGGGGATAGGTGAAAACCGTTACTATCCGTCACTGTCAAATAACAGTGTGGTCGGTGCG
GATAAAGCCAATAACCTGGTTCAGGTGATTCTGTATGGTATCGATCGTACCAACGGCAAA
GGCGAACATATCGTGATGCCTGGCTTTGGTGACGAACTGACCGACAGTCAGATTGCGACC
CTGACTAACTATCTGCGCACCAATTTCCGGCACTAATCCTGCGCCGGTTGATGCCGCACAG
GTGAAAGCGCTGCGGGAAAATAACGTGATGGTTATTCGGGGCTACCTGCTGATTTTGGGC
GGAGTCATCGGTGTCATCATCCTTGTTGCCATCATTATGTACTTCCGTGCGAGAAAAGCT
GCGCGCAACCACGCGGGC

SEQ ID NO.96 - Orf 2420

ATGAAACGATTCTCGCGGGTAAAGCTTACCTTACTGGGGTTGTTGTGCGGCGGTCTGACT
TCACTGGCGGCAAATGCAGCTGACATTGACCAGGCGCTATTGCAACAAGGTGAACAGGTG
GCAACAGCCTCTGACTGTCAGGCTTGTCACACCGCACAGGAGTAAAACCGCATTTCAGT
GGTGGTTATGCAATTGCTTCTCCGATGGGAGCAATATATTCAACCAACATCACTCCGGAT
CCGGCAACAGGTATCGGCAAATACACCGAGCAGCAGTTTATCGAGGCGGTTTCGTATGGT
GTTCCGGGCCGATGGTGCCCAACTGTATCCGGCCATGCCTTATACTTCGTACCGGATGATG
ACTGACAGTGACATCCATGCGCTGTATTACTACTTTATGCATGGTGTGAAACCGGTCGAC
CAGCAGAATACAGAACTCAGCTCTCCTTCCCGTTCAACATGCGTTTTAGCATGAAGTTC
TGGAATCTGCTCTATGCCGACACTAAGACTTTCCAACAGGATCCGCAAAGAGCGCGGAA
TGGAATCGCGGAAATTATCTGGTCAATGGCCTTGCGCACTGTGACACCTGTCATACACCA
CGTGGCTTTATGATGAATGAACAGACCGACCGCTGGCAGGTGCTCCTCTGGGAAGC
TGGTATGCACCGAACATTACTTCAGATAAGGTCAGTGGTATTGGCGGCTGGAGTAACGAT
GAGATAGTTTACGTACCTGAAAACCTGGCCGTGCAGCAGGTAAAAACCGGCGGCTGGCGGG

10/05/2007/04/20/4

10/05/2007/04/20/4

FIGURE 12H

ATGGCAGAAGCCGTGGAACACAGTCTGCAATATCTGCCGGACAGTGATTTACAGGCTATT
GCCACTTATCTGAAGCAAACCACACCGATCCGCACCCCGGGCGAGACTCAGGCGGCATAC
AGCTATGGCTCGTCTTCGACCAATGTTGATGATCAGGTCCGTGGAATGGCACCAAATAAT
GCCCCGTGACTCATTAAACCAGCGGAGCTGCTTTATTTCAGCGGAAGCTGTGCCAGCTGTCAC
CAGCCAGACGGTGCAGGAAGCAAGAATCAGACTTATCCTTCGCTGTTCAATAACACGGCG
ACCGGCATGATTCACCCGCAAAACCTGATTGCAACTATCCTGTTTGGTGTCCAACGTAAC
ACTAAAGACCATCAGGTGCTGATGCCAGGTTTCGGTGCTTCAACCTCCTATGTGGATAGC
CTGACCGATCAACAGATTGCGGATATCAGTAACTATGTACTGCATAATTACGGTAATCCT
GCGGTTACAGTGAAAGCAGGCGATGTGGCGTGGGTTTCGTAAAGGCGGGCATCCGCCGGCA
CTGGTTGCGCTGCAGCCTTATATGATTCCGGCAATTGCGGTTCGGGGTCATTATCATTATC
CTGCTGCTGGTAGCATTTCAGACTTCGTCTAGCCGACGCAAAAGT

Figure 13A

SEQ ID NO. 97 - Orf 3653

ATGTCAGAACAGAACAAAGGGCAGTCTCGCAGGGATTTTTTACTAAAAACGATCACTTTA
GCACCTGCAATGGCTGTCGGAAGCACCGCTATAGGTTCACTTGCTCTTAGCCCGGCAGTA
CAGGCCGCCGATACTCAAACCAGCGGCCCGCAAAGGCCCGGGATTATCAGCCAAACTGG
TTTACCAAAGAAGAGTTTGCATTTATCACTGCTGCTGTGGCGAAACTGATTCCAGCAGAT
TCCCGTGGCCCCGGGAGCCCTTGAGGCGGGTGTACCTGAATACATCGACCGGCAAATGGAT
ACTCCTTATGCCACTGGTTCAAACCTGGTATATGCAAGGCCCGTTTGCCCCCGGATACGCCG
AAAGAACTGGGTTATCAGTTACCGCTGGTGCCGCGGCAAATTTATCGTCTGGGACTGGCC
GATGCAGATAATTTCTGCAAACAACAATATGGTCACGTGTTTGCTGAGCTCAGCGATGAT
CAGCAGGTCACTGCACTGAAAGCTTTTGAATCTGGTCAGGCTAAATTCACCTCAGCTTCCT
GCCACACTGTTTTTTTCTATTTACTACAGAACACCCGCGAAGGTTTCTTCAGCGATCCG
ATCCACGGCGGCAATCAGGGTATGGCTGGCTGGAACTGATTGGTTTCCCTGGTGCCCGG
GCTGACTTTATGGACTGGGTGGAACGTGGTGAACATTATCCGTTCCCACCAGTATCAATT
CGCGGAGAAAGGGCA

SEQ ID NO. 98 - Orf 3689

ATGAAAAATACGCCCCGGAGTAAGGACTCCACCGGCAGACGACTTTTTTTACAACGTTCT
CTATCGCTAATCCCTTTAGTTGCAGCCACAGGTACTCCTTTTGCCACCAGCCAGGCTGCC
GAAAAAAAACCTCCGGCAGTCACTCAGGATTACGTACCGCAATTTTTTGACCCTCAGCAA
TGGGCGTTTATTAATGCCGCAGTTGATCGGCTAATTCCGGAAGATCAGAACGGGGCGGGA
GCTGTCAGTGAAGGTGTTCCGGTCTACATCGATCGTCAGATGGAACCTCCTTATGGTTAC
GGACACCTCTGGTATATGCAGCCCCCTTTCGCATCCCACAGCGACCCGACCCTGGGCTAC
CAGTCCCCTCTGGTGCCTCGTGAGCTTTATCGCCAGGGGATTGCACTCACTGAGCACTAC
TGCCAGCAAACATTTTCATAAGTCGTTTGCTCAACTCACCACCGACCAGCAGGATCAGGTA
TTACAGTTACTCGAAAAGAATACCCTGACGGATAACAATCTGAGTGGTTCGTTATTTTTT
GAGCAACTGCTGGATAACACCAAGGAAGGCTACCTTGAGACCCGGTACATGGAGGCAAT
CAGACTTTGGCTTCATGGAACTGATTGGTTACCCGGGGGCTCGTGAGATTATACCGAC
ACCGTAGCACAGCCAAATGTCCCATACCCGTTGGGCCCTGTGAGTATTTCGGGTAAAAGG
AGTGTC

SEQ ID NO. 99 - Orf 1842

ATGTCAGATAAACCTTCCCATTCCAGGCGTGATTTCTGCTGAAATCACTCACTCTGATC
CCTGCGGTTTCCGTAGGAGGTGCTATTACCAGCGGTATCGCAGGACCGGGCAATGCTCAG
GCGGCCGAAACCTCTGCTACAGCCGCGACAGCGCAGACTCCCTATTCCCCCGTATTTTTT

WO 2005/012547

PCT/US2004/024204

FIGURE 13B

AAACCTGACGAGTGGGCATTTGTGAAAGCAGCCTGCGCCCGGCTGATACCTGCCGATGAT
ATGGGGTCTGGCGCGTTGGAGGCCGGGGTGCCTGAATTTCTCGATCGTCACCTGCAGACC
CCTTATGCCAACGGTTCTGTCTGGTATACCCAGGGGGCCATTTGTTGAGGCAGGACCGGAA
TTTGGCTATCAGGGTCGTAAAACGCTGAGTGAGATCATTCGTTTCAGGGATCCGTGGTGTT
ATCGGCTGGACGCAGAGCAATAAACAGCAGACGTTTCGATGCCCTGACCCATGCAGAACAG
GAAGAAATATTGGTGGCATTGGAAAAAGGCAAGATCCATCTGGAAGAGATGGATGCCAAA
ACCTTCTTCGACTACTTCCTGGGTGAAGTGCGCAATGGCTTCTTTGCTGATCCTTCCTAC
GGAGGCAACAAAGGGATGGTTGGCTGGAAGCTGATCGGCTTCCCGGGGCATGCGTGCCGAT
TACATAGATTTCAATTACCGTCCGCGATAAACCTTATCCGCTCGGACCGGTAGATTTGGCA
GGGAACAGGGGT

SEQ ID NO. 100 - Orf 2037

ATGAAAGAAAATTCTCAACCGCCGGCAGCATCGCGGCGAAAATTTTACAGACAGCCCTG
GCCATTATTCCTTCTACCGCTCTGGCCACCAGTGTCGTGCCTGCTGCTCTGGCGGCCGAA
CAGACCAAAAATCCCACCCGTGATTATGTGCCGGTCTTTTTTAAAGACGATGAGTGGCGG
TTTATTATCGCGGCCACCGATGTGCTGATCCCGGGGGATGAATATGGCCCTGGCGCCGTG
AGTGAAGGAGTTCGGGTGTTTATCGACCGGCAGATGGAAATGCCTTATGGCTACGGTCAG
TTGTGGTACATGAAACCGCCTTTTCAGGAAGGATCTCCACTGCTGGGTACCAGAAAAAC
CTGACTCCACGGGATATCTATCGACGGGGGATCGCCGCCCTGAATAAAGCCTGCCAGACC
ACTTATCAGCATCCGTTTCGCCTCACTGGCGACAGCAGATAAAGTTCAGGTAATGGAAGAT
TTGGAATCCGGAAGCTGGTGACCGAAGACGTTGACGGCAAACCTGTTTTTGCACAGTTA
CTGGAAAATACCAAAGAAGGATACCTTGCCGATCCGATTCACGGGGGCAATCAGACAATG
GCCTCCTGGAAAATGATTGGTTTCCCTGGCGCCCGCGCCGATTACGTTTCAGGTCATGGAT
AATCCCGGAAAACCTTATCTTCCGGGGCCCGGTGAGTATTTCCGGTAAATATGGTGCT

SEQ ID NO. 101 - Orf 2910

ATGAAACAAAGTGGTATCGGGAGGCGTCCGTTTATTATCGGATCTCTGATTGGTATTGCT
TCATTAGGCATGAAGTGTGGTGTAAGTAGTGTTTTTGCAGCTGTCACCTCCCCACTGGAT
GAACTTAACAGTTATCAGCCCGTTTTTTTTTAAACCCGAAGAGTGGCAATTCATCATGGCT
GCCTGCGATCGTCTCATCCCACAGGATGAGGAAGGGCCTGGCGCACTTGAAACACATGTA
CCTGTTTTTATTGATAAACAGATGCTAACACCTTACGGGAAAGGTGAGGACTGGTATATG
GAAGGCCCTTTTAATGCGCATGCCAGCACATTATTTGGCTACCAGTTACCTTTTCCATTG
CAGGTTATGTATCAAAGAGGAATTAAAGCCACCAACAGCTATACCCGCCTCCATTTCAAT
CAGGATTTTGCAGCATTAACTGCGGCACAGCAGGATGCTGTCTTATCAGCACTGGAAGAA

FIGURE 13C

AATAAGATCACTTTCTCAGAGTTTTTCAGAGCCTGACTTATCAGCCTCATATTTCTTTACC
CGGTTGCTGGAAAATACCAAAGAAGGTTATCTGTCTGATCCTAAATATGGTGGAAATAAA
GGCATGGCAGCCTGGGTAATGATCAACTTCCCTGGTGCCCGCGCCAGTTTCCCTACATGG
ATAAAAATTACAAACGTCAAATATCCATTAGGACCGGTAGCTTTGAATGGTGATGTTGCC
CAATCCTCT

SEQ ID NO. 102 - Orf 1951

ATGTCTGATCCTTCATCGAAAGGGATTAGCCGGCGACGGCTGCTCTCTGGCTCTGCCGCA
GGTTTAACCGTTGCCGCGGTAAGCAGTGCTAATGCCACCACTATCACCGGCATCCCTCGC
TGGATGCTGTTTGACCATAACAGCCCCATCACTCCCACCAGCCCCGGCCTTAAGTTCCTG
ACTCAGGAAGAAGCGACTGAAGTGGATGCTATCGTCAGTCAGTTGATCCCTGCCGATGAG
TTGTCGGTGAGCGGTAAAGATGCCGGTTGTACGGTATTTATCGACCGGCAACTGGCTGGC
AGCTATGGCGATGCCAGCCGCAACTATATGCGTGGCCCTTTCCGGGAAGGAACTCCGGCC
CAGGGTGATCAGTCTCCGTTAGTACCCCGCGAGCGTTACCGTCTTGGACTGGCGGGCTTG
AGTGACTATTGCCAGCAAAAATACCAGAACTGTTTCAGCCAACTGGACAGCGCAACCCGG
GATGAAGTGTTGACCGGACTGGAACAGGGAAAAATCAATCTTACCGGCATCAGCGGCAAG
ATGTTTTTTGATCAGGTTCTAACCAACACCATGGAGGGCTTCTTCTCCGATCCGGTGTAT
GGCGGCAACCGCAATATGGTCAGCTGGAAAATGATTGGTTTCCCTGGCGCTCGTTATGAC
TATCGCGACTATCTGACCAAAACCGATCAGAACTGGATTTAGTCCCGATTTCCATCATG
GGCAGCACCGCCTGGAACGCGAAGGTA

SEQ ID NO. 103 - Orf 2634

ATGAAGCGAAGAGAGTTTCTGTCATCAATGGCTGCGTTTGGTGCTGCCTCGGCGATCCCG
CTGACCAATGCTGCCGAAATCTCTGGCGGCCAGCCCTGGCCGCCTGGTCAGGTGAGCCTG
CCTCCGGGGCTTGCCGAGAAAAGGCGGATTACAGTTTTTTACCCGCCATCAACTGGAAACC
GTCGGAGCAATTGCTGAGCGGTTTATTCCCGCCGATGAATTAAGTATCAGTGGTAAAGAG
GCCGGCTGCGCAATTTTTATCGATCGCCAACTGGCAGGGGATTTTGGCCAGGCTGTCACG
GTGTACCGGCTGGGCCGGTTTGTTAAAGGCACTCCTGAGCAGGGGCCACAGTCACCTCTC
ACCCCGGCAGATCAATATCGTCTGGGCCTGAATGCGCTGGACAGCTATTGCCAGCAGCAG
TTTACCATAACTTTACTGAGTTGACCGGTGATCAGCAGGACCAGGTTTTGCAGGGCATG
GAAACCGGGAAAATCAGCCTGGCTGAAAACCTTTGACAGTAAGGTGTTTTTTGAACTGTTA
CTGCAAAACGTCCGTGAAGGTTTTCTGTCCGATCCCCTGTATGGCGGCAACAAAGATATG
GCCAGCTGGAAAATGATTGGTTTTCCCGGTGCCCGTTATGACTTCCGCGATGTTATCGCC
AAAAAAGGCCAAAAATTAAACATTATTCCTACCAGCCTGATTGATAACAACCTT

FIGURE 13D

SEQ ID NO. 104 - Orf 764

ATGCTTTTGCAAAAAACACCACGCGCCGCAAATTCCTGCTCGGTTCGCTGATGGCTTTG
CCTGAACTTACCAGCTATAAACCAGGCTTTTTTCACCGCTGACGAATGGCAGTTTATTCTG
GCTGCGACCGACCGCATTATTCCTGCGGGCGGACCGGGTAAAGCCCCGGGCGCGCTGGAA
ACTAATGTGCCGATATTTATCGACCAGCAACTCCATGATGAGCATTTCGGTAAGGAAATC
TACATGGAAGGGCCGTTTAACCCGCATGCCCCGGCCACTATGGGGTATCAGGTGCCTCTC
TATCCACAACAGATTTATCAGACCGGTATCCGGCTGACTAATCAGTGGAGCCAGCAAAAC
CTGCAGAAACCTTTCCATCAGTTATCGGAAGCAGATAAAGACAAGGTGCTGACGGGATTA
CAGAAAAACACTCTCGACTTCGCAGCCCTGGGTGAAAACACCCTGAAGGGCTCGTTGTTC
TTCAGTCAGTTGCTCGGAGAAACCAACACGGTTATCTCGCCGACCCGATGTATGGCGGC
AATAAAGGGATGAAAGCGTGGATTGCGATTGGTTTTCCCGGGGCCCGCGCCAGTTATCTG
GAATGGGTAAAACAACATAATGTGAAGTATCCCCTCGGGCCGGTCAGCCTGCTGGGCGAG
ACTGCG

SEQ ID NO. 105 - Orf 3051

ATGCAACGTCGTAAATTTATCAAGACCGGATTAATCCTGGCAGGGACCGGAACTGCAGCA
TCAGTATTTAAACCAGCGGGTGCTGCGGCGCGCGATAATACTGAATGGCGGAAAACCTG
TGGAAGCTAAAGAAACACCACCGCCGACTCCGGCTGATCCAACCAAACGTCTCTATCTG
ACCGAACAGGAATATGCCAGATCACCGCGATTTTTTAACCGGCTGATCCCTGCAGATGAA
CTGACTGTCAGCGCCTCCGATGCGGGGCTGTGTTGTTTTATCGATAACCAGTTAGCCGGA
AATTATGGTAAAGCCAGCTGGCGCTATAATGTTGGCCCGTTTGAAAATGGTACGCCTTCC
CAGGGTAACCAGCAGCCTTACACTCCGGCTCAGATTTACCGTATTGGTTTGGCCGAAATA
GAAAAAGACTGTCAGAGTAAATTTCAAATCCTTCAGCGAACTAACTAATGATCAGCAG
GATAAATATCTGGAACAGATGGAAGCCGACCAGATTAAATACCCTACCCTGTCATCCAAA
GATGTATTTAGTCAGTTCTTATCCAATGTACAGGAAGGTTTTCTTGCCGACCCGATATAT
GGTGGTAACCGCAATATGATTGGCTGGAAAATGATTGGTTTCCCGGGAGCACGTTATGAC
TATCGTGATTATGCCCCACTGAAAGGAACTAAATTAATATCGAACCGGTCAGCATTATT
CAACTCCTGAAAGCA

SEQ ID NO. 106 - Orf 1748

GTGAAGCGCAGGCGTTTTTTGGCTTCTCTCGGAGTATTGCTTATCAGCACTGCTCTGAAA
GTAAAGCAAAGATTATTTCCGGCGGTATGCCGTGGGTGCTGCATGCTGTTAAGCCACCG

11/02/2013 11:11:11

11/02/2013 11:11:11

FIGURE 13E

CAACCAGTAGTCGCGGGGGAATGGCAGTTTTTTACACCAGAAGAAGTGGCGATAATTGAA
GCTATTGCTGACCGGATAATCCCTCAGGATGAACTGAGTATTGGCGGAAAAGAAGCGGGT
TGCGCATTATTTCTCGACCGCCAGTTAGCGGGAGATTACGGCAAAGCAGTCAGTATATAT
CGTCTTGGACCGTTTATTCAGAATGGTTTACCGGAGGCGGGCCCGCAATATAAAGATGTC
CCTGCAGAACGTTACCGGTTGGGTCTGGCGTCAGTAAATGAAATCAGCCAGGCCAAATAC
AATGGTAAAAAGTTCAATGAAATCAGTGAAGAACAACAGGATGATTTACTGGGTAAAATC
GAATCGGGAGTATTACCACTCACCGGAGTCGACGGTAAGTTATTTTTTGATCAGTTGGTC
ATAAATATGCGTGAAGGATTTTTTGCCGATCCATTATATGGCGGAAATAAAGATATGGCT
GGCTGGAAAATGCTCGGTTTTCCCGGCGCGCAATATGATTTCCGTGATGTGATTGATAAA
CGAGGCGAAGAATTGAATATCAAGCCGGTCAGCATGGTAACTAACAACGATCAATCT

Figure 14A

SEQ ID NO. 107 - Orf 3819

ATGACAGCAAATAATCGCCATCCTTCCGGGGTTTCACGCCGTCGTCTGCTGCAGGGTATG
GGCATTCTGTGCGGTTGCCGGGTTATGTGGTTCACGTGTTTCCTTCTTTTCGTGCAGCAGCA
GCAGAACTGCAGGACAGTGGTTTTATCCCGTTATCTGAATTTCTGGTTAATCGCCGGGTG
AACCCAATTCTGGCTCAGCGTTACTACGATGCATTGCATCGCCATGATGAAAAATTTGAT
CAGAAGCTGGCATTGCTTAAACAGGATATTCAGCCAGGAAAGTATCAAAACATTGATGAT
TTTCTTCAAAAAAATGCCGTCGGAACAGATTTACGACAGGCAGCAGGTCAGGTCATTTCT
GCCTGGTACACCGGAGTGGTAGGCAACGACGAGAACTGGAAGTTATCGCATATGCAGAC
GCGATGATGTATGTGCCACCAAGTGGCGTTCTTGTGGTTCCAACCTATGGCAGTGGCCCG
ATTTCTGGGCTGCCGTTGACAATAAACCCGCACACCAGGGGGCCGGCTGTA

SEQ ID NO. 108 - Orf 2447

ATGAAACTCACAGATACTATTTTCGACAGATCGCCGAAACTGATTAAGTCATTATCGTTG
CTGACCGTATTCTCTGTGAGTGGGTTACGTCTGGTGACCTGTCCGGCATTGCTGGCGGT
TTGCCTGCCAGTGCGGATTTTCATGAATTCTCAACCTTTGTTATTGGCCGGCCAGTAGAT
CCTGTTTTATCAGGCCGTTATTTTGCCGCATTGCAGGCTGCAGACGGACATTTCAATCAA
CAACTGAATCAGGCGATGGTTGCCAGTGTCCCGTTTCGCAGTCAGGGGATTGATACGATG
CTGGCATCACTCCCTCACGACAGTGACATCTTTAATAACCTTAAAAAATCACCTCAGCC
TGGTATCTGGGAATCGTCGGCGAAGGCGCCGGGGCGACTCTGATCGCCTTCCATGACGCA
CTGATGTTCCAGCCAACCCGTGAATACGTTTTTGTTCCCGGTTATGGCGGGGGCCCTGAC
AGTTGGGTCTCACTTAAACATCCCGACTTACTGAGCGAAGATACCGAGCAGGAACAGAAA
AATGGC

SEQ ID NO. 109 - Orf 3676

ATGAAAATGAAATTATTCGGGATGACTCTCCTGCTGAATACAATTTGTCCCGCAGAAAG
GTGCTGCTTGGTGGTCTTATTTTATTAGGCAGTAGTTATCTGGGCCCATCGCTTCCGGCC
TGGGCAGATACGTAAATGACCAGGCTACCATCGACCAGTTTATGCAGTTATCACAATTA
CTTGTTAATCATCAGCTTGATCCGGTAACAGGGCAGCGTCTGGCTGCTGCGATGATCAGT
GGCAACATGATTACACGGCAACAGATAACCAGTCTGCTGGCTGTAGCTCAGGCCCGTCAG
GCAAAAGTGGTTGAGGATTTTTCTCAGATATTCCACAGGGTGAGCTGAAAAATGCGGCT
CTCAGCATTATCTCTGCCTGGTATAAAGGTGTACTGATTGATGCACCCGGGGCTGAGGTG
TTTGCTTATGAAAAGCCTTAATGTATCAGCCGACTATCGATGTGATGACCATTCCGACC
TATGCCATTACCGGCCCTAATGGCTGGAGCTCCCATGCAGCTCCGCTGGCCGATATGCCG
GACTTC

WO 2005/012547

PCT/US2004/024204

FIGURE 14B

SEQ ID NO. 110 - Orf 1219

ATGATTGATATGTTAAATATGATATCGCGCAGGCGCATATTGCAGGGGATGGGGGCATTG
GCCGCCACAACATTACTTCCGTCTGGTATTTTACCGGCATTGCGGGATACTCCTGCAAAC
AGCGACTTTAACGATATATCCAGGTTGTTAACCGGCCGCAATACTTTATCGGCTGAATTT
AGTAGCGCGCTTTTTTCTGCCTTCACGAAAATTGATAGCCGCTTCCCACAGCAACTGGCT
CGCCTGAAACAGTGGATCACTGCTAATTCCGTCCCGGCAGCTGACCTGCAGAAACGTCTG
ACCGCCGACAGCTCTGTGCGGGATCTGGCCGGATTACCGGCACTGATTCTGACCGGCTGG
TATCTGGGTATTGCCGGCAGCGGCGACAAAGCTGTCTGCGTCACCTATGTCGATGCTCTG
GCAAATCAGGAAGTTGCATCGGTGCTGAACCCGCCTACCTACGCCTATGGCGCCTATGGC
AGTTGGGCCACAAAACCTTTT

SEQ ID NO. 111 - Orf 1957

ATGAATAACCATAATGCACCTGAAACACAGCCTGAGCTGAGTGAAGAGGGGTTGCGTCGC
CGAAAGCTGTTCCGACAGACGGGCGGGTTGGTGGCTTCCTTCGCCATTGGCTCCGCAATA
GCTGGCAGCACACTCAGTAATGGTGCCAATGCAGCAACCACTTCTGCCGGGCCTGATACT
CAGACCCTGAATCAGTTTATGAAAACCTCCCGTCTTCTTACCGGCCATCAGAACCTCGAT
CTTACCCTGGGTCAACGTCTGTACGTGGCATTGAGCGAGAAGGACCCACAGTTCATTACC
CAGCTGTCAGCCTTGAATCAGTGGATTGCCGATAAACAACCGGCAGATGTAGAGGCCCTT
GACAGCCAGTTGTCAGGACAGCCGCTGCATGCCCTGATGATGTCGGTGATTAAAGGCTGG
TATCTCGGAGTCATTGATGACAGTCACCATGCCAAAGTCTATGCCTATCAGAATGCCCTG
ATGTACCAGGTACCGCGTGATGGCATGGTCATCCCTACCTATGCGCATAACGGACCGGAC
TACTGGACTGCAGATCCACCACCGGTCGATCGACTTCTGAACTTC

SEQ ID NO.112 - Orf 3043

ATGAATAAGGCTACACCTGTCAGCCCCGGTGAAAGACGACGCTTTATTAAGTTGTTAGCA
GCATCAACGGTGGCGGGAACAGTCAGCAGTTTATTGCCGGGCCAGATAGCCTGGGCAATT
GATGCCGGGCAGCCGGCGGTGCGCCGGTTTTCCGGCATTATGACGGTATCTGAAATTATT
TGCGGCTATCCGACTCTGGATAATGCACTGGGCAAACGTATTTTCAGCCTGATCAGTGCT
GAGCACGGTGATGCTTCTCAGAGTATTGCCGAGCTGCAAAAGCAACTGAATGCAGATATG
TCCTCTGCCGAAATGCAGGCGGCGTTGAAAACACTGGATACCCCGGCACAACAGCTGTTT
AGTGAAATTTTGCGCGGCTGGCAAATTGGTATTGTCGGTAGCGGTAAGCAATCACAGGTC
GTGGCTTATGAGTACGCACTGATGTACGCGCCGATTCAGATGTTGTCGTCCTGCCGACG
TTTGCCCGTGGTGAACCCCACTACTGGGCATACCCTCCTGTGATTAAGACCGGAAAGCTG

FIGURE 14C

SEQ ID NO. 113 - Orf 1982

ATGAAATTTGTTATAGATCAGGAATCTGATACAGGGGAAATCTCTGCATCACGCCGTAGT
TTCCTGATAAAAATAACTGCATTATTAGCCTCATTACCTTGATTCCGGCACATGCTGTT
ATTACCACTCCAGCTGACGTTGGTGCATCAGTGATATCGCAGTTACAGACAACCGCTCAA
TTCTTAACCGAAAGCCAGCAGGATCCTCAACTGATTATCCGTGCAGCTAACGCCCTGCTA
AAAGTTAACAGTAATTTTGCCGGTGATCTGCAACAGCTTTCTTCATTGATTGCAGACAAT
CACATAGCCAACTTAAAAGACCTCAAACCTCAAATCTTTTTGAGGGCAAACCACAGCAA
ACCGCGAAAGACATTTTGTCTGCCTTATATCTTGGCTATGCCGGAACACCGGTGATGTTG
TCATCGGAAGATAATGTCGTGTTTGTGCTTATGCCAGGCACGCACGTATCAGCTCACC
AAAGATTTACCCCCGGTCCCCAGCTACTCCCGCTGGAAAAGTGGCTACTGGGCGCATTG
CCGGCAGGCGTT

SEQ ID NO. 114 - Orf 3398

ATGAATCTTACACGCCGCCGGTTGCTGACCGGCTCGGCGGGGCTGATAGTGGCTGGCGTA
TTGTCGCAGACTCTGTCAGGCCGCTATGCGCTGGCCAGTCCGCCACTGGCTTCCGCAGTC
GCCCCCTCCGCCGGATTTAACACATTATCGGTACTGATTACCGGTCAGGATAAGCCCGAT
GCCCTGCTGGCGCAGCGCCTGTACAGCTGGTTAGCAGCCCATACTTCCGGTCTGGACAGT
CAGTTGGAGACGCTGAGCTCACTGCTGCAACAACACTCTGATGCTAATGGCAGCACCTG
CTTAGCCTGATGAAATCGCAGCCAGAAAATATTAATACTCTATCAGTCACTGGTGTCT
GGCTGGTATCTCGGTGTTGTCGGGCCACTGCCGCGTCCGGACTGCATCGCCTTCGAAAAT
ATTGTCAGCTACCAACTGCTCAAACAATCTGTGTTACCGCCAAGCTACGCGCCGGGGCCAA
CCAGGATTCTGGGTGCAGCCACCTGCGGGGAGAGTACATGTC

SEQ ID NO. 115 - Orf 2418

ATGAAGCAAATATTTGAGCAAAGTCATACCGATCTACCGGAAAATGGAACCGGTTCCAGT
CGCAGAGGATTTATTAAGTCCGCTCTGGTATTAAGTCCAGTGGTCTGGTCGCGTCTCTG
CCATTGCGTAGTTTCGCCAGCAGTGTGGTTCATGGTGGCGATACCACTCAGGACTTTATC
AGTGTTTCGCAGGCAATCACCGAACACAAACATATCAACCCACAGTTAGCCGCTCATTT
CTGAGTGCGTTTATCAAAGGGATAATCAGTTCAGCAGCAAATTACCCGACTTGCGCAG
CTCTACCAGACGGGTGATACAGCTATTGTATTTAAAAACAAAGCGGTAGCCGCCGGGCTT
GGCGATTTTCTGCAGCAGATCCTGACCGCCTGGTATACCGGAACGATTGGTGATGACTAC
AAAGGCACTCTGGTCGCTTACAAAGAAGCGCTGATGTACGACACCGTGAGCGATGGCTTA
GTGGTCCCGACCTATTGCGGCAATGGCCCGCTTTGGTGGACAGTGCCGGTCCCCGACCCA
CTCGATCCTGAACTGATCAACAACCTG

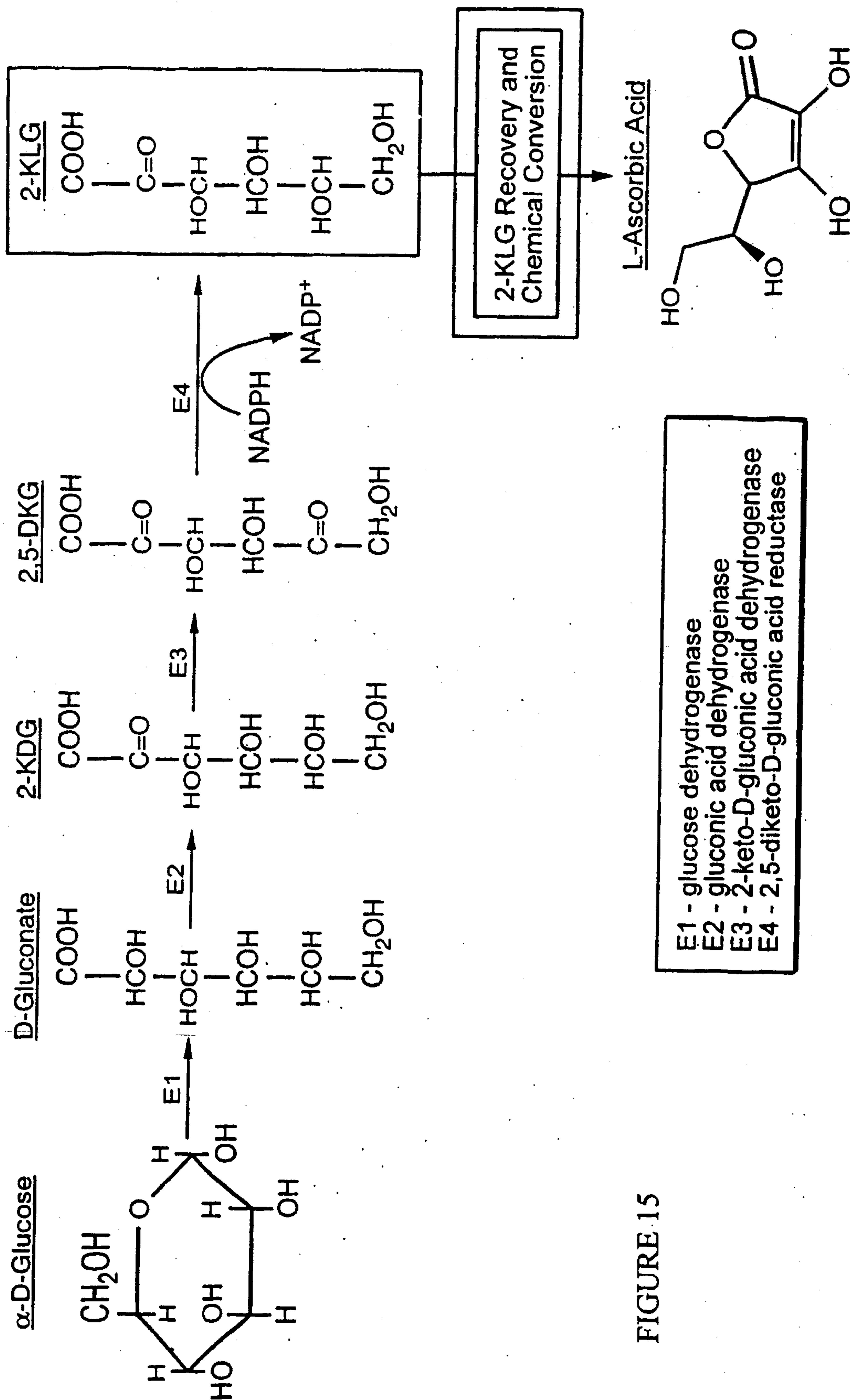


FIGURE 15

Heme-binding site 1

40	G	E	Y	L	A	R	L	G	D	C	M	A	C	H	T	T	S	G	R	P	orf-766.PRO
49	G	Q	Y	L	A	T	A	G	D	C	A	A	C	H	T	Q	P	G	G	K	orf-1840.PRO
35	G	R	Y	I	A	Q	L	G	D	C	T	A	C	H	T	E	P	G	H	P	orf-1953.PRO
34	G	E	Y	L	A	T	A	G	D	C	V	A	C	H	T	A	-	S	G	P	orf-2035.PRO
55	G	E	Y	I	A	V	A	G	D	C	V	A	C	H	T	A	P	G	S	K	orf-2632.PRO
39	G	A	Y	L	A	R	A	G	D	C	T	A	C	H	T	E	A	G	G	K	orf-2908.PRO
54	G	E	Y	I	A	R	L	S	D	C	T	A	C	H	T	A	E	G	G	Q	orf-3053.PRO
31	G	E	Y	L	A	R	A	G	D	C	V	A	C	H	T	S	E	G	G	Q	orf-3651.PRO
38	G	E	Y	L	A	R	A	G	D	C	V	A	C	H	T	A	E	G	G	K	orf-3687.PRO
53	G	K	Y	L	A	T	A	G	D	C	V	A	C	H	T	A	P	G	G	K	orf-1750.PRO
41	G	K	Y	L	S	V	A	A	D	C	G	A	C	H	N	S	P	T	S	G	orf-3820.PRO
58	G	H	Y	L	A	I	A	A	D	C	A	A	C	H	T	D	P	Q	T	K	orf-1956.PRO
34	G	E	Y	L	A	R	A	G	D	C	T	A	C	H	T	A	A	G	G	A	orf-1984.PRO
37	G	E	Q	V	A	T	A	S	D	C	Q	A	C	H	T	A	P	G	S	K	orf-2420.pro
37	G	K	Y	L	S	V	A	A	D	C	S	A	C	H	D	S	P	D	H	-	orf-2446.PRO
60	G	E	Y	L	A	K	A	A	D	C	E	V	C	H	T	A	T	G	G	Q	orf-3041.PRO
33	G	E	Y	L	A	T	A	S	D	C	T	A	C	H	T	A	P	G	G	P	orf-3400.PRO
38	G	E	Y	L	A	T	A	S	D	C	A	A	C	H	T	D	K	G	G	-	orf-3674.PRO
36	G	Q	Y	L	S	V	A	G	D	C	A	A	C	H	T	T	S	G	G	K	orf-1221.pro



FIGURE 16A

Heme-binding site 2

193	L	G	H	C	S	S	C	H	T	P	R	G	V	A	M	E	E	K	A	L	orf-766.PRO
205	L	G	H	C	G	T	C	H	T	P	R	A	G	T	L	Q	E	K	A	L	orf-1840.PRO
187	L	G	H	C	S	T	C	H	T	P	K	N	M	-	-	-	-	-	-	L	orf-1953.PRO
186	L	E	H	C	G	A	C	H	T	P	R	G	I	A	F	Q	E	K	A	S	orf-2035.PRO
208	L	E	H	C	A	A	C	H	T	P	K	N	M	-	-	-	-	-	-	L	orf-2632.PRO
198	L	G	H	C	S	S	C	H	T	P	R	G	I	G	M	Q	E	K	A	L	orf-2908.PRO
205	P	G	H	C	A	S	C	H	T	A	K	N	M	-	-	-	-	-	-	L	orf-3053.PRO
185	L	G	H	C	G	A	C	H	T	P	R	S	I	T	M	Q	E	K	A	L	orf-3651.PRO
191	L	E	H	C	G	A	C	H	T	P	R	G	I	A	F	Q	E	K	A	L	orf-3687.PRO
205	L	G	H	C	T	S	C	H	T	P	K	N	M	-	-	-	-	-	-	L	orf-1750.PRO
195	L	E	H	C	D	A	C	H	T	P	R	N	F	L	M	-	-	-	-	-	orf-3820.PRO
212	L	E	H	C	G	E	C	H	T	P	R	N	T	L	M	-	-	-	-	-	orf-1956.PRO
187	L	A	H	C	S	T	C	H	S	P	R	N	Q	M	M	-	-	-	-	-	orf-1984.PRO
191	L	A	H	C	D	T	C	H	T	P	R	G	F	M	M	-	-	-	-	-	orf-2420.pro
190	L	A	H	C	D	T	C	H	T	P	R	N	A	L	M	-	-	-	-	-	orf-2446.PRO
212	L	G	H	C	E	E	C	H	T	P	R	N	L	A	M	-	-	-	-	-	orf-3041.PRO
185	L	A	H	C	T	T	C	H	T	P	R	N	L	T	M	-	-	-	-	-	orf-3400.PRO
191	A	A	H	C	S	T	C	H	T	P	R	G	P	L	M	-	-	-	-	-	orf-3674.PRO
189	L	A	H	C	S	T	C	H	T	P	R	N	A	L	M	-	-	-	-	-	orf-1221.pro



Heme-binding site 3

308	D	L	T	E	G	Q	T	L	Y	L	N	N	C	G	A	C	H	F	V	N	G	L	D	A	A	-	orf-766.PRO
326	N	P	T	P	G	A	Q	L	Y	V	D	N	C	S	A	C	H	Q	T	S	G	K	E	V	Q	-	orf-1840.PRO
298	Q	M	V	N	G	K	A	I	Y	A	D	R	C	S	A	C	H	V	S	Q	G	K	E	V	S	-	orf-1953.PRO
305	L	S	Q	T	G	A	E	E	Y	M	B	N	C	S	A	C	H	R	L	D	G	K	G	Y	A	-	orf-2035.PRO
319	L	M	Q	Q	G	A	N	V	Y	Q	A	N	C	S	A	C	H	N	S	D	G	R	G	I	P	-	orf-2632.PRO
317	N	L	N	S	G	Q	M	L	Y	L	N	N	C	S	A	C	H	L	T	D	G	G	G	A	K	-	orf-2908.PRO
316	Q	M	V	S	G	K	K	V	F	E	S	Q	C	S	A	C	H	V	S	D	G	A	G	I	F	-	orf-3053.PRO
307	D	Q	R	T	G	A	A	E	Y	V	D	S	C	S	A	C	H	K	T	D	G	S	G	Y	T	-	orf-3651.PRO
309	V	S	A	T	G	A	Q	V	Y	L	D	N	C	S	A	C	H	R	S	D	G	K	G	Y	D	-	orf-3687.PRO
316	V	M	A	R	G	Q	S	I	Y	Q	A	N	C	S	A	C	H	Q	S	D	G	A	G	V	F	-	orf-1750.PRO
321	-	P	D	A	G	W	V	V	F	S	N	T	S	C	A	N	C	H	Q	A	N	G	E	G	S	-	orf-3820.PRO
345	D	D	I	T	G	K	A	L	Y	D	T	T	C	A	S	C	H	Q	S	S	G	A	G	T	T	D	orf-1956.PRO
319	S	T	T	D	G	A	V	L	F	D	S	S	C	A	S	C	H	Q	P	D	G	A	G	S	K	N	orf-1984.PRO
321	S	L	T	S	G	A	A	L	F	S	G	S	C	A	S	C	H	T	P	S	G	A	G	S	Y	S	orf-2420.pro
319	G	P	M	S	G	A	I	L	Y	S	G	N	C	A	S	C	H	T	P	S	G	A	G	S	Y	S	orf-2446.PRO
338	S	A	P	S	G	A	V	L	F	N	G	N	C	A	S	C	H	T	P	S	G	A	G	S	Y	S	orf-3041.PRO
312	Q	S	Q	P	G	A	V	V	F	E	G	N	C	S	T	C	H	E	A	G	G	Q	G	I	G	E	orf-3400.PRO
320	A	K	Q	Q	G	L	A	L	F	Q	G	N	C	A	S	C	H	E	A	G	G	Q	G	S	R	D	orf-3674.PRO
313	D	-	-	-	G	A	K	L	Y	E	A	H	C	S	T	C	H	Q	A	S	G	Q	G	S	N	G	orf-1221.pro



FIGURE 16B