



## (51) International Patent Classification:

*C12N 9/00* (2006.01)      *C07H 21/00* (2006.01)  
*C12N 15/29* (2006.01)      *C12N 15/31* (2006.01)  
*A01H 1/00* (2006.01)      *A01H 5/00* (2006.01)

## (21) International Application Number:

PCT/US2015/019058

## (22) International Filing Date:

5 March 2015 (05.03.2015)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

61/949,082      6 March 2014 (06.03.2014)      US

(71) Applicants: **SALK INSTITUTE FOR BIOLOGICAL STUDIES** [US/US]; 10010 North Torrey Pines Road, La Jolla, California 92037 (US). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Oakland, California 94607 (US).

(72) Inventors: **NOEL, Joseph**; 14428 Rock Rose, San Diego, California 92127 (US). **WOODS, Kate**; 10010 North Torrey Pines Road, La Jolla, California 92037 (US). **BOWMAN, Marianne**; 10010 North Torrey Pines Road, La Jolla, California 92037 (US). **LOUIE, Gordon**; 10010 North Torrey Pines Road, La Jolla, California 92037 (US). **DA SILVA, Nancy Anderson**; 4912 Rain Tree Lane, Irvine, California 92612 (US). **CARDENAS, Javier Pascual**; 742 Turtle Crest Drive, Irvine, California 92603 (US).

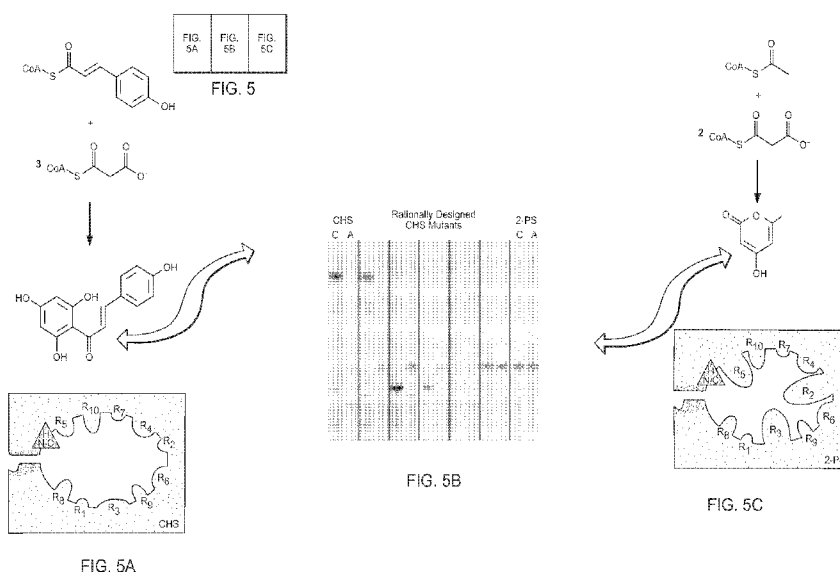
(74) Agents: **HUNTER-ENSOR, Melissa** et al.; Saul Ewing LLP., Center Square West, 1500 Market Street, Philadelphia, Pennsylvania 19102 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on next page]

## (54) Title: POLYKETIDE SYNTHASE VARIANTS AND USES THEREOF



(57) Abstract: The present invention relates to methods for producing polyketide synthase variants, and for altering the activity and/or substrate specificity of putative native and variant polyketide synthases. The present invention further relates to compositions comprising said polyketide synthase variants, compounds prepared using said polyketide synthase variants, and uses of said polyketide synthase variants. In one embodiment, said polyketide synthase variant is 2-pyrone synthase.



DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT,  
LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE,  
SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

**Declarations under Rule 4.17:**

— *of inventorship (Rule 4.17(iv))*

## **POLYKETIDE SYNTHASE VARIANTS AND USES THEREOF**

### **CROSS REFERENCE TO RELATED APPLICATIONS**

5 This application claims benefit of U.S. Provisional Application Serial No. 61/949,082, filed March 6, 2014, the contents of which are incorporated herein by reference.

### **STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH**

10 This invention was made with government support under Grant Nos. EEC-0813570 and MCB-0645794 awarded by the National Science Foundation. The government has certain rights in the invention.

### **SEQUENCE LISTING**

15 The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on March 5, 2015, is named 365550.1003WO1\_SL.txt and is 335,357 bytes in size.

### **BACKGROUND**

25 There has been great interest in utilizing either naturally occurring enzymes or engineered enzymes to prepare commercially useful chemical compounds. 2-Pyrone (also termed 2PY) is an attractive target for large-scale commercial production because it serves as a starting material for the synthesis of at least three distinct commodity chemicals. Through chemical catalysis, 2-Pyrone can be converted to sorbic acid, a preservative used widely in dairy-based foods. Moreover, further chemical modification yields the enone, 3-pentene-2-one, which is useful for chemical synthesis, and the diene 1,3-pentadiene, which is useful as a plasticizer. Additionally, 30 2-pyrone itself may be used as a food additive, as a phytochemical or as part of a phytochemical mix.

2-Pyrone synthase (also termed 2-PS) catalyzes the synthesis of 2-Pyrone using acetyl-CoA as one of the starting materials. 2-Pyrone synthase or other 2-Polyketide synthase variants

having altered capabilities could expand the enzymatic toolbox available for rational enzyme-based chemical synthesis. More specifically, the discovery of 2-polyketide synthase variants is required to provide novel non-native enzymes having new synthetic capabilities and/or enhanced enzymatic activity.

5

## SUMMARY OF THE INVENTION

As described below, the present invention provides polyketide synthase variants having altered enzymatic activity, host cells (e.g., yeast, plant, algae, bacterial, mammalian, insect) comprising such variants and *in vitro* and *in vivo* methods of using such enzymes for the  
10 production of desired chemical compounds, including polyketide, pyrone and lactone products.

In one aspect, the invention provides an isolated nucleic acid molecule encoding a polyketide synthase variant having an altered amino acid residue corresponding to amino acid positions C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.

15 In another aspect, the invention provides a vector containing the isolated nucleic acid molecule of any aspect described herein.

In still another aspect, the invention provides a cell containing the isolated nucleic acid molecule and/or vector of any aspect described herein.

In yet another aspect, the invention provides a plant containing the isolated nucleic acid  
20 molecule and/or vector of any aspect described herein. In various embodiments, the nucleic acid molecule is expressed in a glandular tricone.

In one aspect, the invention provides an isolated polyketide synthase polypeptide having an altered amino acid residue corresponding to amino acid positions C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.

25 In another aspect, the invention provides a method of increasing the activity of a polyketide synthase, the method involving introducing an altered amino acid residue at a position corresponding to C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.

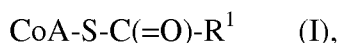
In another aspect, the invention provides an isolated nucleic acid molecule encoding a polyketide synthase variant comprising an altered amino acid residue corresponding to amino acid positions C35, C65, C89, C135, C195, C346, C372, T137, I201, L202, and/or L268 of 2-

pyrone synthase. In various embodiments, the altered amino acid residue is any one or more of C35S, C65S, C89S, C135S, C195S, C346A, C372S, L202F, L268F, L268M, and T137F. In particular embodiments, the altered amino acid residue is any one or more of C35S, C65S, C89S, C135S, C195S, C346A, C372S, L202F, L268F, L268M, and T137F. In one embodiment, the encoded polyketide synthase variant comprises C35SC372S.

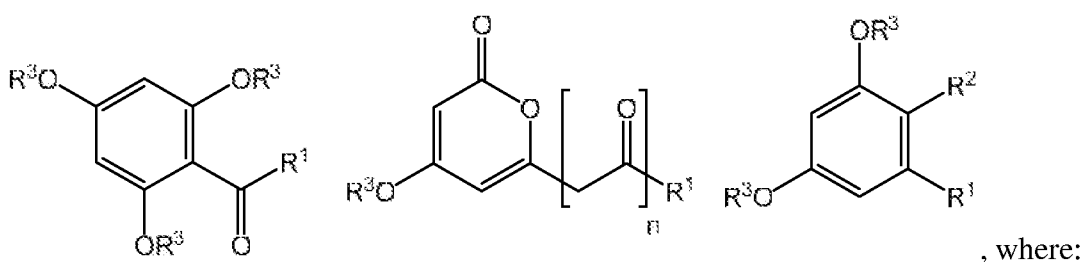
In another aspect, the invention features an isolated polyketide synthase polypeptide comprising an altered amino acid residue corresponding to amino acid positions C35, C65, C89, C135, C195, C346, C372, T137, I201, L202, and/or L268 of 2-pyrone synthase. In particular embodiments, the altered amino acid residue is any one or more of C35S, C65S, C89S, C135S, C195S, C346A, C372S, T137F, L202F, L268F, and L268M. In one embodiment, the polyketide synthase polypeptide comprises C35SC372S.

In another aspect, the invention features a method of increasing the activity of a polyketide synthase, the method comprising introducing an altered amino acid residue at a position corresponding to C35, C65, C89, C135, C195, C346, C372, T137, I201, L202, and/or L268 of 2-pyrone synthase. In one embodiment, the altered amino acid residue is selected from the group consisting of C35S, C65S, C89S, C135S, C195S, C346A, C372S, T137F, L202F, L268F, and L268M.

In still another aspect, the invention provides a method of preparing a cyclic compound, or a salt, solvate or derivative thereof, the method involving contacting the isolated polyketide synthase of any aspect delineated herein with a substrate of formula (I), or a salt, solvate or derivative thereof:



where the cyclic compound is selected from the group consisting of:



R<sup>1</sup> is selected from the group consisting of C<sub>1</sub>-C<sub>20</sub> alkyl, C<sub>1</sub>-C<sub>20</sub> cycloalkyl, aryl-(C<sub>1</sub>-C<sub>20</sub> alkyl), carboxy-(C<sub>1</sub>-C<sub>20</sub> alkyl), aryl, heterocyclyl, heteroaryl, aryl-CH=CH-, heterocyclyl-CH=CH- and heteroaryl-CH=CH-,

$R^2$  is selected from the group consisting of H, carboxy,  $C(=O)O(C_1-C_6)alkyl$ , cyano, - $C(=O)NH_2$ , - $C(=O)NH(C_1-C_6)alkyl$ , and - $C(=O)N((C_1-C_6)alkyl)_2$ ;

each occurrence of  $R^3$  is independently selected from the group consisting of H,  $C_1-C_{20}$  alkyl, aryl- $(C_1-C_{20} alkyl)$ ,  $C_1-C_{20}$  acyl, aroyl, and heteroaroyl; and,

n is 0, 1, 2 or 3;

where each alkyl, cycloalkyl, aryl, heterocyclyl or heteroaryl group is independently optionally substituted with at least one substituent selected from the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro, monoalkylamino, dialkylamino, carboxy,  $C(=O)O(C_1-C_6)alkyl$ , trifluoromethyl, cyano, - $C(=O)NH_2$ , - $C(=O)NH(C_1-C_6)alkyl$ , - $C(=O)N((C_1-C_6)alkyl)_2$ , - $SO_2NH_2$ , and - $C(=NH)NH_2$ .

In various embodiments,  $R^1$  is  $C_1-C_{20}$  alkyl, carboxy- $(C_1-C_{20} alkyl)$ , aryl, heteroaryl, aryl- $CH=CH-$ , or heteroaryl- $CH=CH-$ . In particular embodiments,  $R^1$  is  $C_1-C_{20}$  alkyl, carboxy- $(C_1-C_{20} alkyl)$ , aryl, or aryl- $CH=CH-$ . In various embodiments, each occurrence of aryl comprises phenyl.

In still another aspect, the invention provides a method of preparing a polyketide using a cell containing the isolated nucleic acid molecule and/or vector of any aspect described herein.

In another aspect, the invention provides a kit containing an isolated polyketide synthase, a nucleic acid molecule, a vector or a cell of any aspect described herein.

In various embodiments of any of the aspects delineated herein, the polyketide synthase variant is 2-pyrone synthase. In various embodiments the altered amino acid residue is selected from the group consisting of C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M. In particular embodiments, the altered amino acid residue is C35S. In certain embodiments, the polyketide synthase variant has a combination of altered amino acid residues that is C35S, T137F; C35S, T137L; C35S, I201F; C35S, I201L; C35S, L202F; C35S, L202M; C35S, L268F; C35S, L268M; C35S; C35S, C65S; C35S, C89S; C35S, C195S; C35S, C346A; C35S, C372A; C35S, C372S; C35S, C89S, C372S; C35S, C65S, C89S, C372S; C35S, C89S, C195S, C346A, C372S; and C35S, C65S, C89S, C195S, C346A, or C372S. In specific embodiments, the polyketide synthase polypeptide has a combination of altered amino acid residues that is C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M.

In various embodiments of any of the aspects delineated herein, the activity and/or stability of the polyketide synthase is increased relative to a reference polypeptide (e.g., wild-

type 2-pyrone synthase). In particular embodiments the  $k_{cat}$  of the polyketide synthase is increased and/or  $K_m$  of the synthase is decreased relative to a wild-type 2-pyrone synthase polypeptide. In various embodiments, the activity of the polyketide synthase involves production of a compound that is one or more of a pyrone, a chromone, a triacetic lactone, a polyhydroxynaphthalene, a phloroglucinol, a resorcinol, a resorcinol acid, SEK 4, SEK4b, and  
5 aloesone, including combinations thereof. In a particular embodiment, the lactone is triacetic lactone.

In various embodiments of any of the aspects delineated herein, the cell is a yeast, plant, algae, bacterial, mammalian, or insect cell. In various embodiments, the variant polyketide  
10 synthase is expressed in a yeast, plant, bacteria, mammal, or insect cell. In various embodiments, the cell is *in vitro* or *in vivo*.

Other aspects, embodiments, advantages, and features of the present invention will become apparent from the following specification.

## 15 Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed.  
20 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

By “polyketide synthase (PS) polypeptide” is meant a polypeptide or fragment thereof  
25 having at least about 85% amino acid identity to UniProt Accession No. P48391 and having polyketide synthase activity. Polyketide synthase activity includes, but is not limited to, Claisen condensing carbon-carbon bond formation, aldol condensing carbon-carbon bond formation, lactonizing carbon-oxygen bonding formation, decarboxylation and/or (thio)ester cleavage. In other embodiments, a polyketide synthase polypeptide has at least about 85%, 90, 95, 96, 97, 98,  
30 99, or 100% identity to a wild-type polyketide synthase polypeptide. Accordingly, the term

polyketide synthase polypeptide encompasses polyketide synthase variants. Sequences useful in the methods of the invention are provided at Figures 9-13.

Polyketide synthases include any one of a family of enzymes that catalyze the formation of polyketide compounds. Polyketide synthases are generally homodimers, with each monomer being enzymatically active. An exemplary polyketide synthase is the *Gerbera hybrida* 2-pyrone synthase (2-PS) (also termed Gh2PS), which amino acid sequence is shown below.

```

1 mgsyssddve vireagraqg latilaigta tppncvaqad yadyyfrvfk sehmvdlkek
61 fkricektai kkrylalted ylqenptmce fmapslnarq dlvtgtgvpml gkeaavkaid
10 121 ewglpkskit hlifcttagv dmpgadyqlv kllglspsvk rymlyqqgca aggtvlrlak
181 dlaennkgsr vlivcseita ilfhgpnneh ldslvaqalf gdgaaalivg sgphlaverp
241 ifeivstdqt ilpdtekamk lhlreggltf qlhrdvplmv aknienaaek alsplgitdw
301 nsvfwmvhpq graildqver klnlkedklr asrhvalsey nlisacvlfi idevrkrsma
361 egksttgegl dcvlfgfgp gmtvetvvlr svrvtaavan gn (SEQ ID NO: 1)

```

By “polyketide synthase (PS) nucleic acid molecule” is meant a polynucleotide encoding a polyketide synthase polypeptide or fragment thereof. An exemplary PS nucleic acid molecule sequence is provided under NCBI Accession No. Z38097, and is shown below, with the protein-encoding region spanning bases 163 to 1371 in bold.

```

1 aaaaggccta ctcaagcctt gaaattctct tttcttttct tttcattccc ttccctcaaa
61 ttataaactt acctttctgt ttctttcaaa gaatttagct gcctcaaacg aagatcttca
121 tatctcattt gttaggatat acaaacatca atctcgagta aaatggggatc ttactcatcc
181 gatgatgtgg aggtgattcg tgaggccgga cgggcacaag gtttagccac gattcttgcc
25 241 attggcactg ctactcctcc caattgcgtc gctcaagctg attatgcaga ctattatatt
301 cgtgtcacta agagcgaaca tatggttgat cttaaagaga aatttaaacg catttgtgag
361 aaaacagcga taaagaaacg atacctagcc ctcaccgaag actatctgca agagaaccca
421 acaatgtgtg agttcatggc tccatcctta aacgctcgac aagacctagt ggtcaccggc
481 gtcccaatgc ttggcaaaga agccgcagtc aaggccattg atgaatgggg actaccaaaa
30 541 tccaagatca cccacctcat cttctgcacc accgctggcg ttgacatgcc cgggtgtgac
601 tatcaactcg tcaaactcct tggctctctc ccttcagtca aacgctatat gttgtaccaa
661 cagggatgtg ccgccggcgg cacagtcctc cggctagcca aggaccttgc tgaaaacaac
721 aagggtcac gaggccttat cgtctgctcc gagatcactg ctatcttatt ccatggaccc

```

781 aatgagaacc accttgactc actcgtcgct caagctttat tcggagacgg agctgcagca  
 841 ctcatgtgtg gttcaggccc tcacttgccc gtagaacggc caatattcga gatcgtgtca  
 901 actgatcaaa caatcttgcc ggacactgag aaggcaatga agttacactt gagagagggga  
 961 gggttgacgt ttcagttgca tagagatgta cccttgatgg tcgcaaagaa catagagAAC  
 5 1021 gcagcggaga aagcgttgct tccactaggg ataactgatt ggaactcagt tttctggatg  
 1081 gtgcaccag gtggtcgagc catattggac caggtggagc gaaaactaaa ccttaaggaa  
 1141 gataagttaa gggctagcag gcatgtgctt agtgaatacg gaaacctgat tagcgttgt  
 1201 gtgttggtca tcattgacga ggtgaggaag agatctatgg cggaggga gagtacaacc  
 1261 ggtgaagggt tggattgcgg tggtttgttt ggatttgac cgggtatgac tgttgagact  
 10 1321 gttgttcttc gtagcgtccg cgttactgct gcggttgcca atggaaactg atcactgttg  
 1381 tttgcaaat attacttttt actacggtat gtttccttgt ttatgagttt gtcattcacc  
 1441 tatgataata ggtctgtat ttttcttgtt tatgatatta ttttctcaaa gatgatgtaa  
 1501 gttggcaatt aaataaagat tgtttttctt atgaataata taagattaca ttttc (SEQ  
 ID NO: 2)

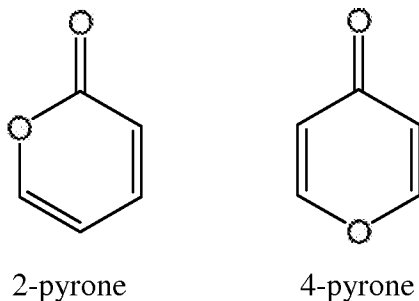
15

20

As used herein, the term “substrate” refers to a Coenzyme-A (CoA) thioester that acts as an enzymatic substrate for a wild-type polyketide synthase or a polyketide synthase variant contemplated within the invention. In certain embodiments, a substrate comprises acetyl-CoA, malonyl-CoA, coumaroyl-CoA, hexamoyl-CoA, a synthetic CoA thioester, ACP or NAC

thioesters and the like. In certain embodiments, the term “substrate” refers to CoA thioesters found in natural systems, as well as CoA thioesters that are chemically synthesized and not found in natural systems.

As used herein the term “pyrone” refers to a class of cyclic chemical compounds having the molecular formula  $C_5H_4O_2$  and comprising an unsaturated six membered ring comprising one oxygen atom and a ketone functional group. Two isomers denoted as 2-pyrone (2PY) and 4-pyrone (4-PY) are exemplified herein. As used herein, the term “pyrone” also refers to a compound comprising a 2-pyrone moiety or a 4-pyrone moiety, and derivatives thereof.



By “complex” is meant a chemical association of two or more molecules. Complexes may include a network of weak electrostatic bonds that maintain the association of the molecules. Other types of interactions, such as covalent bond, ionic bond, hydrogen bond, hydrophobic interaction, or van der Waals interactions, may be present instead of or in addition to electrostatic bonds between members of a complex.

By “specifically binds” is meant a compound or antibody that recognizes and binds a polypeptide of the invention, but which does not substantially recognize and bind other molecules in a sample.

By “computer modeling” is meant the application of a computational program to determine one or more of the following: the location and binding proximity of a ligand to a binding moiety, the occupied space of a bound ligand, the amount of complementary contact surface between a binding moiety and a ligand, the deformation energy of binding of a given ligand to a binding moiety, and some estimate of hydrogen bonding strength, van der Waals interaction, hydrophobic interaction, and/or electrostatic interaction energies between ligand and binding moiety. Computer modeling can also provide comparisons between the features of a model system and a candidate compound. For example, a computer modeling experiment can compare a pharmacophore model of the invention with a candidate compound to assess the fit of the candidate compound with the model. Examples of techniques useful in the above evaluations include: quantum mechanics, molecular mechanics, molecular dynamics, Monte Carlo sampling,

systematic searches and distance geometry methods. Further descriptions of computer modeling programs are provided elsewhere herein.

As used herein, “molecular replacement” refers to generating a preliminary model of a polyketide synthase whose structural coordinates are unknown, by orienting and positioning a molecule whose structural coordinates are known within the unit cell of the unknown crystal so as best to account for the observed diffraction pattern of the unknown crystal. Phases can then be calculated from this model and combined with the observed amplitudes to give an approximate Fourier synthesis of the structure whose coordinates are unknown. This in turn can be subject to any of the several forms of refinement to provide a final, accurate structure of the unknown crystal (Lattman, E., 1985, in *Methods in Enzymology*, 11 5.55-77; Rossmann, MG., ed., “The Molecular Replacement Method” 1972, *Int. Sci. Rev. Ser.*, No. 13, Gordon & Breach, New York). Using structure coordinates of the polyketide synthase provided herein (see *e.g.*, PDB Accession Numbers) molecular replacement may be used to determine the structural coordinates of a crystalline mutant, homolog, or a different crystal form of polyketide synthase.

By “decreases” is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By “increase” is meant a positive alteration of at least 10%, 25%, 50%, 75%, or 100%.

In one embodiment, the presence of a mutation increases the activity of an enzyme of the invention by at least about 5%, 10%, 15%, 20%, 25%, or 50%.

The term “detect” refers to identifying the presence, absence, level, or concentration of an analyte.

By “foreign” or “heterologous” is meant a gene encoding a protein whose exact amino acid sequence is not normally found in the host cell. Exemplary host cells include yeast, plant, algae, bacterial, mammalian, and insect cells.

By “fragment” is meant a portion of a polypeptide or nucleic acid molecule. This portion contains at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

As used herein, “-carbon” refers to the chiral carbon atom found in an amino acid residue. Typically, four substituents will be covalently bound to said -carbon including an amine group, a carboxylic acid group, a hydrogen atom, and an R-group.

As used herein, “R-group” refers to the substituent attached to the  $\alpha$ -carbon of an amino acid residue. An R-group is an important determinant of the overall chemical character of an amino acid. There are twenty natural R-groups found in proteins, which make up the twenty naturally occurring amino acids.

5 By “substantially identical” is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). In one embodiment, such a sequence is at least 60%, 80%, 85%, 90%, 95% or even 100% identical at the amino acid level or nucleic acid to the sequence used  
10 for comparison (e.g., wild-type reference sequence).

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar  
15 sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a  
20 probability score between  $e^{-3}$  and  $e^{-100}$  indicating a closely related sequence.

As used herein, “active site” refers to a site in a synthase defined by amino acid residues that interact with substrate and facilitate a biosynthetic reaction that allows one or more products to be produced. Active site amino acids for 2-pyrone synthase include L202 and L268. The three-dimensional space position of an  $\alpha$ -carbon at the active site of a synthase and of R-groups  
25 associated therewith can be determined using techniques such as three-dimensional modeling, X-ray crystallography, and/or techniques associated therewith. Active sites can be specified by a set of amino acid residues. Other residues can play a role in substrate specificity and enzyme activity by modulating size, shape, charge, and the like of the active site. In addition, residues external to the active site may also modulate the specificity and/or activity of the enzyme.

30 As used herein, “hydrophilic amino acid” and “hydrophilic R-group” includes any naturally occurring or unnatural amino acid that is relatively soluble in water and/or has high

affinity for water. Examples of naturally occurring hydrophilic amino acids include serine, threonine, tyrosine, asparagine, glutamine, cysteine, and the like.

As used herein, “hydrophobic amino acid” and “hydrophobic R-group” includes any naturally occurring or unnatural amino acid that is relatively insoluble in water and/or has low  
5 affinity for water. Examples of naturally occurring hydrophobic amino acids are alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, methionine, and the like.

As used herein, “naturally occurring amino acid” and “naturally occurring R-group” includes L-isomers of the twenty amino acids naturally occurring in proteins. Naturally occurring amino acids are glycine, alanine, valine, leucine, isoleucine, serine, methionine,  
10 threonine, phenylalanine, tyrosine, tryptophan, cysteine, proline, histidine, aspartic acid, asparagine, glutamic acid, glutamine, arginine, and lysine. Unless specially indicated, all amino acids referred to in this application are in the L-form.

As used herein, “negatively charged amino acid” and “negatively charged R-group” includes any naturally occurring or unnatural amino acid having a side chain which is negatively  
15 charged under normal physiological conditions. Examples of negatively charged, naturally occurring amino acids include aspartic acid, glutamic acid, and the like.

As used herein, “positively charged amino acid” and “positively charged R-group” includes any naturally occurring or unnatural amino acid having a side chain which is positively charged under normal physiological conditions. Examples of positively charged, naturally  
20 occurring amino acids include arginine, lysine, histidine, and the like.

As used herein, “unnatural amino acid” and “unnatural R-group” includes amino acids that are not naturally found in proteins. Examples of unnatural amino acids included herein are racemic mixtures of selenocysteine and selenomethionine. In addition, unnatural amino acids include the D- or L- forms of, for example, nor-leucine, para-nitrophenylalanine,  
25 homophenylalanine, para-fluorophenylalanine, 3-amino-2-benzylpropionic acid, homoarginines, D-phenylalanine, and the like. Unnatural amino acids and peptides including such amino acids are described in U.S. Pat. Nos. 6,566,330 and 6,555,522.

By “variant” as is meant a polynucleotide or polypeptide sequence that differs from a wild-type or reference sequence by one or more nucleotides or one or more amino acids.  
30 Exemplary 2-pyrone synthase variants include those having an amino acid change at any one or more of T137, I201, L202, L268, C35, C65, C89, C195, C346, and C372.

In one embodiment, exemplary 2-pyrone synthase variants include the following combinations of the following amino acid substitutions: C35S, T137F; C35S, T137L; C35S, I201F; C35S, I201L; C35S, L202F; C35S, L202M; C35S, L268F; C35S, L268M; C35S; C35S, C65S; C35S, C89S; C35S, C195S; C35S, C346A; C35S, C372A; C35S, C372S; C35S, C89S, C372S; C35S, C65S, C89S, C372S; C35S, C89S, C195S, C346A, C372S; and C35S, C65S, C89S, C195S, C346A, and C372S.

In another embodiment, the enzyme comprises at least C35S, C65S, C89S, C195S, C346A, and C372S. This background might be useful for generating active site variants.

As used herein, “isolated” refers to a molecule that is substantially free of other elements present in its natural environment. For instance, an isolated protein is substantially free of cellular material or other proteins from the cell or tissue source from which it is derived. The term “isolated” also refers to preparations where the isolated protein is sufficiently pure to be administered as a pharmaceutical composition, or at least 70-80% (w/w) pure, more preferably, at least 80-90% (w/w) pure, even more preferably, 90-95% pure; and, most preferably, at least 95%, 96%, 97%, 98%, 99%, or 100% (w/w) pure.

As used herein, “native” or “native polypeptide” refers to proteins that occur or are produced in nature (*e.g.*, wild-type 2-pyrone synthase).

As used herein, “mutant” or “mutated polypeptide” refers to a polypeptide containing amino acid residues that have been substituted or modified with respect to a wild type polypeptide (*e.g.*, polyketone synthase).

By a “substantially pure polypeptide” is meant a polypeptide (*e.g.*, a pyrone synthase) which has been separated from components which naturally accompany it or accompany it *in vitro*. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. In one embodiment, the preparation is at least about 75%, 85%, 90%, 95%, or 99%, by weight, pure. A substantially pure synthase polypeptide may be obtained, for example, by extraction from a natural source; by expression of a recombinant nucleic acid encoding an synthase polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method (*e.g.*, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis).

The term “promoter” as used herein is defined as a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a polynucleotide sequence.

As used herein, the term “promoter/regulatory sequence” means a nucleic acid sequence which is required for expression of a gene product operably linked to the promoter/regulatory sequence. In some instances, this sequence may be the core promoter sequence and in other instances, this sequence may also include an enhancer sequence and other regulatory elements which are required for expression of the gene product. The promoter/regulatory sequence may, for example, be one which expresses the gene product in a tissue specific manner.

By “reference” is meant a standard or control condition.

A “reference sequence” is a defined sequence used as a basis for sequence comparison.

As used herein, “space group” refers to the arrangement of symmetry elements within a crystal.

As used herein, “structure coordinates” refers to Cartesian coordinates (x, y, and z positions) derived from mathematical equations involving Fourier synthesis as determined from patterns obtained via diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of a polyketide synthase molecule in crystal form. Diffraction data are used to calculate electron density maps of repeating protein units in the crystal (unit cell). Electron density maps are used to establish the positions of individual atoms within a crystal's unit cell. The term “crystal structure coordinates” refers to mathematical coordinates derived from mathematical equations related to the patterns obtained on diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of a synthase polypeptide (*e.g.*, a chalcone synthase protein molecule) in crystal form. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are used to establish the positions of the individual atoms within the unit cell of the crystal. The crystal structure coordinates of a synthase can be obtained from crystals and can also be obtained by means of computational analysis.

By “three-dimensional model” is meant a three-dimensional representation of a molecule's structure. Computer modeling may be used to generate such a model in conjunction with structural data. These data could include x-ray crystallographic data, nuclear magnetic

resonance data, electron microscopy data, or any other source of experimental or theoretical data useful for generating a model of a molecule or complex of molecules.

By “unit cell” is meant the fundamental repeating unit of a crystal.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having “substantial identity” to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By “hybridize” is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (*e.g.*, a gene described herein), or portions thereof, under various conditions of stringency. (See, *e.g.*, Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, *e.g.*, formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, *e.g.*, sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 g/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 g/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature.

As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS.

Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al.* (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

As used herein, the term “alkyl,” by itself or as part of another substituent means, unless otherwise stated, a straight or branched chain hydrocarbon having the number of carbon atoms designated (*i.e.*, C<sub>1</sub>-C<sub>10</sub> means one to ten carbon atoms) and includes straight, branched chain, or cyclic substituent groups. Examples include methyl, ethyl, propyl, isopropyl, butyl, isobutyl, tert-butyl, pentyl, neopentyl, hexyl, and cyclopropylmethyl. Most preferred is (C<sub>1</sub>-C<sub>6</sub>)alkyl, such as, but not limited to, ethyl, methyl, isopropyl, isobutyl, n-pentyl, n-hexyl and cyclopropylmethyl.

As used herein, the term “cycloalkyl,” by itself or as part of another substituent means, unless otherwise stated, a cyclic chain hydrocarbon having the number of carbon atoms designated (*i.e.*, C<sub>3</sub>-C<sub>6</sub> means a cyclic group comprising a ring group consisting of three to six carbon atoms) and includes straight, branched chain or cyclic substituent groups. Examples include cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, and cyclooctyl. Most preferred is (C<sub>3</sub>-C<sub>6</sub>)cycloalkyl, such as, but not limited to, cyclopropyl, cyclobutyl, cyclopentyl and cyclohexyl.

As used herein, the term “substituted alkyl,” “substituted cycloalkyl,” “substituted alkenyl” or “substituted alkynyl” means alkyl, cycloalkyl, alkenyl or alkynyl, as defined above, substituted by one, two or three substituents selected from the group consisting of halogen, -OH, alkoxy, tetrahydro-2-H-pyran-2-yl, -NH<sub>2</sub>, -N(CH<sub>3</sub>)<sub>2</sub>, (1-methyl-imidazol-2-yl), pyridin-2-yl, pyridin-3-yl, pyridin-4-yl, -C(=O)OH, trifluoromethyl, -C≡N, -C(=O)O(C<sub>1</sub>-C<sub>4</sub>)alkyl, -C(=O)NH<sub>2</sub>, -C(=O)NH(C<sub>1</sub>-C<sub>4</sub>)alkyl, -C(=O)N((C<sub>1</sub>-C<sub>4</sub>)alkyl)<sub>2</sub>, -SO<sub>2</sub>NH<sub>2</sub>, -C(=NH)NH<sub>2</sub>, and -NO<sub>2</sub>, preferably containing one or two substituents selected from halogen, -OH, alkoxy, -NH<sub>2</sub>, trifluoromethyl, -N(CH<sub>3</sub>)<sub>2</sub>, and -C(=O)OH, more preferably selected from halogen, alkoxy and -OH. Examples of substituted alkyls include, but are not limited to, 2,2-difluoropropyl, 2-carboxycyclopentyl and 3-chloropropyl.

As used herein, the term “aryl,” employed alone or in combination with other terms, means, unless otherwise stated, a carbocyclic aromatic system containing one or more rings (typically one, two or three rings) wherein such rings may be attached together in a pendent manner, such as a biphenyl, or may be fused, such as naphthalene. Examples include phenyl, anthracenyl, and naphthyl. Preferred are phenyl and naphthyl, most preferred is phenyl.

As used herein, the term “heterocycle” or “heterocyclyl” or “heterocyclic” by itself or as part of another substituent means, unless otherwise stated, an unsubstituted or substituted, stable, mono- or multi-cyclic heterocyclic ring system that consists of carbon atoms and at least one heteroatom selected from the group consisting of N, O, and S, and wherein the nitrogen and sulfur heteroatoms may be optionally oxidized, and the nitrogen atom may be optionally quaternized. The heterocyclic system may be attached, unless otherwise stated, at any heteroatom or carbon atom that affords a stable structure. A heterocycle may be aromatic or non-aromatic in nature. In one embodiment, the heterocycle is a heteroaryl.

As used herein, the term “heteroaryl” or “heteroaromatic” refers to a heterocycle having aromatic character. A polycyclic heteroaryl may include one or more rings that are partially saturated. Examples include tetrahydroquinoline and 2,3-dihydrobenzofuryl.

Examples of non-aromatic heterocycles include monocyclic groups such as aziridine, oxirane, thiirane, azetidine, oxetane, thietane, pyrrolidine, pyrroline, imidazoline, pyrazolidine, dioxolane, sulfolane, 2,3-dihydrofuran, 2,5-dihydrofuran, tetrahydrofuran, thiophane, piperidine, 1,2,3,6-tetrahydropyridine, 1,4-dihydropyridine, piperazine, morpholine, thiomorpholine, pyran, 2,3-dihydropyran, tetrahydropyran, 1,4-dioxane, 1,3-dioxane, homopiperazine, homopiperidine,

1,3-dioxepane, 4,7-dihydro-1,3-dioxepin and hexamethyleneoxide.

Examples of heteroaryl groups include pyridyl, pyrazinyl, pyrimidinyl (such as, but not limited to, 2- and 4-pyrimidinyl), pyridazinyl, thienyl, furyl, pyrrolyl, imidazolyl, thiazolyl, oxazolyl, pyrazolyl, isothiazolyl, 1,2,3-triazolyl, 1,2,4-triazolyl, 1,3,4-triazolyl, tetrazolyl, 1,2,3-thiadiazolyl, 1,2,3-oxadiazolyl, 1,3,4-thiadiazolyl and 1,3,4-oxadiazolyl.

As used herein in a chemical context, the term “substituted” means that an atom or group of atoms has replaced hydrogen as the substituent attached to another group.

For a cyclic group, the term “substituted” as applied to the ring(s) of these groups refers to any level of substitution, namely mono-, di-, tri-, tetra-, or penta-substitution, where such substitution is permitted. The substituents are independently selected, and substitution may be at any chemically accessible position. In one embodiment, the substituents vary in number between one and four. In another embodiment, the substituents vary in number between one and three. In yet another embodiment, the substituents vary in number between one and two. In yet another embodiment, the substituents are independently selected from the group consisting of C<sub>1-6</sub> alkyl, -OH, C<sub>1-6</sub> alkoxy, halo, amino, acetamido and nitro. As used herein, where a substituent is an alkyl or alkoxy group, the carbon chain may be branched, straight or cyclic, with straight being preferred.

Unless specifically stated or obvious from context, as used herein, the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

Any compounds, compositions, or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

Unless specifically stated or obvious from context, as used herein, the term “or” is understood to be inclusive. Unless specifically stated or obvious from context, as used herein,

the terms “a”, “an”, and “the” are understood to be singular or plural. Thus, for example, reference to “an amino acid substitution” includes reference to more than one amino acid substitution.

The term “including” is used herein to mean, and is used interchangeably with, the phrase “including but not limited to.”

As used herein, the terms “comprises,” “comprising,” “containing,” “having” and the like can have the meaning ascribed to them in U.S. Patent law and can mean “includes,” “including,” and the like; “consisting essentially of” or “consists essentially of” likewise has the meaning ascribed in U.S. Patent law and the term is open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited is not changed by the presence of more than that which is recited, but excludes prior art embodiments.

Other features and advantages of the invention will be apparent from the following description of the desirable embodiments thereof, and from the claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B are graphs showing a comparison of triacetic lactone (TAL) production by various *S. cerevisiae* strains expressing wild-type or variant *G. hybrida* 2-pyrone synthase (2-PS). Figure 1A shows variants bearing amino-acid substitutions at sites within the enzyme’s active site. Figure 1B shows variants bearing amino-acid substitutions at sites external to the enzyme’s active site. For each variant, the amino-acid substitution(s) with respect to the wild-type (WT) 2-PS sequence is(are) indicated. For the measurement of triacetic lactone titers, yeast strains were cultured in selective SDC(A) medium (1% dextrose, 0.67% yeast nitrogen base, 0.5% Bacto casamino acids, 0.5% ammonium sulfate and 100 mg/L adenine), aliquots of the cultures were measured turbidimetrically for cell density and then centrifuged to pellet cell material, and TAL in the supernatant was quantitated by HPLC analysis with separation by reversed-phase chromatography on a C18 column.

Figure 2 provides a schematic overview of polyketide biosynthesis highlighting the potential for the generation of a large number of biologically useful cyclic compounds. Diversity in the final polyketide products (examples are shown at the lower right) can originate from variation in the starter molecular-group (lower left box: shown are acetyl, octanoyl, malonyl, methyl-anthraniloyl, isopentanoyl, and coumaroyl); the identity and sequence variant of the type-

III polyketide-synthase enzyme (represented by the ribbon diagram of a PKS homodimer); the number of iterations of chain extension; and the mechanism of cyclization (which is governed sterically by the shape of the enzyme's active-site cavity, as represented pictorially at the top right).

Figure 3 provides a schematic depiction of an intermediate of polyketide biosynthesis highlighting reactive chemical groups. Bottom figure: a type-III PKS enzyme catalyzes iterative chain extension via decarboxylative condensations with malonyl-CoA to generate a repetitive and flexible linear oligomer carried on a catalytic cysteine. Top figure: within the extended polyketide intermediate, the carbonyl groups of the individual building-blocks can undergo keto-enol tautomerization (dashed orange box), and these tautomeric states and reactive-group proximity dictate intramolecular covalent-bond formation with consequent cyclization of the polyketide oligomer. In some embodiments, the activation barrier is raised to alter polyketide intermediate activity.

Figure 4 depicts catalyzed reactions and active-site cavity structures of two polyketide-synthase enzymes, chalcone synthase (left) and 2-pyrone synthase (right). The top panels show molecular surface representations of the enzymes bound to their respective substrates (the CHS/naringenin complex derives from a crystal structure, which guided in part the modeling of methylpyrone into the 2-PS structure). (Phe 220 and Phe 270 in 2-PS, and Phe 215 and Phe 265 in CHS were removed for clarity in these depictions.) Highlighted in color on each molecular surface are the catalytic cysteine residue (red), three residues at which targeted amino-acid substitutions confer 2-PS-like activity on CHS (green), and a site at which amino-acid substitution has no effect on these enzymes' product specificity (blue). The bottom panels show schematic representations of the active-site cavities of the two enzymes, highlighting the surrounding R-groups that sterically define the volume of each active-site cavity.

Figure 5 depicts steric engineering to alter the substrate specificity of chalcone synthase from coumaryl CoA to acetyl CoA. The radiogram shows thin-layer chromatographic analyses of the reaction products generated by CHS or 2-PS enzymes with the substrates radiolabeled malonyl-CoA and either p-coumaroyl-CoA (C) or acetyl-CoA (A) as the starter molecule: leftmost lane, wild-type CHS; rightmost lane, wild-type 2-PS; middle lanes, several CHS variants bearing amino-acid substitution(s) at sites as described in Figure 4. The mobilities of naringenin (formed from the spontaneous cyclization of chalcone) and methylpyrone are

indicated by blue and red (respectively) arrows. The variant chalcone-synthases that have been sterically engineered to accept acetyl CoA catalyze only two chain extensions and thereby produce triacetic acid lactone.

Figure 6 provides a schematic representation of the 2-pyrone synthase active-site cavity. Highlighted are the constellation of amino-acid residue R-groups that sterically define the volume of the active-site cavity, and the observed alterations in enzyme properties resulting from the modification of these residues.

Figure 7 is a graph showing increased triacetic lactone (TAL) production through 2-PS expression in *S. cerevisiae* strains with engineered disruption of the endogenous yeast genes encoding various proteases or metabolic enzymes. The knockouts affect genes involved in intracellular (vacuolar) protein degradation (PEP4, an aspartyl protease; and PRB, proteinase B); gluconeogenesis (PYC2, a pyruvate carboxylase); and regulation of lipid metabolism (NTE1, serine esterase). The maximal level of triacetic acid lactone production obtained was 1.32 g per liter of culture, with a yeast strain ( $\Delta pyc2 \Delta nte1 \Delta pep4 \Delta prb1$ ) with four genes disrupted. Figure 8 depicts potential chemical intermediates and end-products identified and synthesized from triacetic acid lactone.

Figure 8 depicts chemical conversions of the starting reagent triacetic acid lactone that generate exemplary intermediates or end-products useful, for example, in chemical syntheses.

Figure 9 is a non-limiting list of 2-pyrone synthase variants, including the number of amino acid substitutions in each one relative to the wild-type amino acid sequence and the specific amino acid substitution identified by position.

Figure 10 is a list of amino acid sequences (SEQ ID NOS 3-63, respectively, in order of appearance) of the 2-pyrone synthase variants.

Figure 11 is an alignment of the amino acid sequences (SEQ ID NOS 1 and 3-63, respectively, in order of appearance) of the 2-pyrone synthase variants.

Figure 12 is a list of nucleic acid sequences (SEQ ID NOS 64-124, respectively, in order of appearance) of the 2-pyrone synthase variants.

Figure 13 is an alignment of the nucleic acid sequences (SEQ ID NOS 125 and 64-124, respectively, in order of appearance) of the 2-pyrone synthase variants

Figure 14 includes a graph and a table. The graph compares triacetic lactone (TAL) titers (g/L) for 2-pyrone synthase (2-PS) variants C35A, C35S, L268F, and C35SL268F relative to the

wildtype *G. hybrida* 2-pyrone synthase (WT). Bars represent mean values  $\pm$  one standard deviation (n=6 independent experiments).

Figure 15 includes a graph and a table. The graph compares triacetic lactone (TAL) titers (g/L) for 2-pyrone synthase (2-PS) variants with active site mutations at positions L202, L268, I343, M259, T137, and I201 in combination with the C35S mutation. Bars represent mean values  $\pm$  one standard deviation (n=6 independent experiments).

Figure 16 provides a graph and a table. The graph compares triacetic lactone titers (g/L) for 2-Pyrone synthase variants with the C35S mutation and external cysteine mutations (to serine or alanine) at positions C65, C89, C135, C195, C346, and C372S. Bars represent mean values  $\pm$  one standard deviation (n=6 independent experiments).

Figure 17 provides a graph and a table. The graph compares triacetic lactone (TAL) titers (g/L) for 2-Pyrone synthase variants with either the native C35 or modified C35S, and incorporation of additional cysteine modifications. Bars represent mean values  $\pm$  one standard deviation (n=6 independent experiments).

Figure 18 provides two graphs. The graph compares triacetic lactone (TAL) levels (g/L and g/g dry cell weight (DCW)) for BYt (light gray), and protease-knockout strains (BYt $\Delta$ *prb1* (checkered), BYt $\Delta$ *pep4* $\Delta$ *prb1* (hashed), and BJ5464 (light gray-hashed)) with wild-type 2-Pyrone synthase or variants C35S or C35SC372S. Bars represent mean values  $\pm$  with one standard deviation (n=6 biological replicates).

## DETAILED DESCRIPTION OF THE INVENTION

As described below, the present invention provides polyketide synthase variants having altered enzymatic activity, host cells (e.g., yeast, plant, algae, bacterial, mammalian, insect) comprising such variants and *in vitro* and *in vivo* methods of using such enzymes for the production of desired chemical compounds, including polyketide, pyrone and lactone products.

The invention is based, at least in part, on the discovery of 2-pyrone synthase variants having increased enzyme activity. More specifically, as reported in detail below, the invention provides polyketide synthase variants having increased triacetic acid lactone production relative to a wild-type polyketide synthase enzyme. Such enzymes are useful, for example, for enhancing the enzymatic synthesis of polyketide, pyrone and lactone products relative to wild-type enzyme.

Accordingly, the invention provides polyketide synthase variants proteins having increased enzyme activity, polynucleotides encoding the polyketide synthase variants, host cells (e.g., yeast, algae, plant, bacterial, mammalian, insect) comprising such polynucleotides), and methods of using these host cells for the production of desired compounds (e.g., polyketide, pyrone and lactone products).

### Polyketide Synthases

Polyketides are a class of compounds that includes a broad range of antibiotics, immunosuppressants and anticancer agents. Polyketides encompass molecules with extremely rich bioactivities, including antibiotics (e.g., tetracyclines and erythromycin), anti-cancer agents (e.g., daunomycin), immunosuppressants (e.g., FK506 and rapamycin), and veterinary products (e.g., monensin). Many polyketides (produced by polyketide synthases) are valuable as therapeutic agents. Polyketide synthases are multifunctional enzymes that catalyze the biosynthesis of a variety of carbon chains differing in length and patterns of functionality and cyclization (Figure 2).

In particular, the compound 2-pyrone is an attractive target for large-scale commercial production because it serves as a starting material for the synthesis of a variety of commodity chemicals. Through chemical catalysis, 2-pyrone can be converted to sorbic acid, a preservative used widely in dairy-based foods. Moreover, further chemical modification yields the enone, 3-pentene-2-one, which is useful for chemical synthesis, and the diene, 1,3-pentadiene, which is useful as a plasticizer. In addition, 2-pyrone itself may be of value as a food additive, as a phytochemical, or as part of a phytochemical mix.

2-Pyrone synthase (also termed 2-PS) is used to catalyze the synthesis of 2-pyrone. As described herein, the enzyme-engineering methods described herein with 2-pyrone synthase constitute a generally useful approach that can be applied to virtually any 2-pyrone synthase-type (type-III polyketide synthase) enzyme for enhancing overall small molecule production in a yeast metabolic-engineering host or other suitable hosts. Atomic resolution x-ray crystal structures were found to be important in the rational engineering of the enzyme to accept a variety of CoA-thioester starter molecules, the number of decarboxylative acetyl additions and the termination chemistry used to offload useful products (Claisen and aldol condensations and lactonization) (Figure 3). Moreover, the molecular structure of 2-pyrone synthase is useful for characterizing

variants that greatly improve *in vivo* stability and production through the replacement of key Cys residues that are either externally exposed to the bulk solvent or buried. Replacing Cys residues defined by the x-ray structures with non-oxidizable amino acids has the potential to increase productivity of the polyketide synthase enzymes *in vitro* and *in vivo*, which is useful for biocatalysis and bioengineering. Thus, the effects of particular Cys residues was determined by creating and characterizing such variants in 2-pyrone synthase.

A 2-pyrone synthase polypeptide of the invention encompasses wild-type polyketides (*e.g.*, a *G. hybrida* 2-pyrone synthase amino acid sequence), or 2-pyrone synthase variants having one or more mutations (*e.g.*, those described in Figure 10 or otherwise delineated herein) relative to the wild-type sequence. Mutations include deletions and additions of amino acid residues, and substitutions of one amino acid residue for another. Exemplary substitutions include: C35S (where Cys at position 35 of a wild type 2-pyrone synthase is changed to Ser), I343F, I343M, I201M, L202F, L268M, T137F, T137L, I201F, I201L, L202F, L202M, L268F, L268M, C35A, C65A, C346S, C135A, C65S, C89S, C195S, C346A, C372A, C372S mutants, variants and conservative substitutions thereof comprising L- or D-amino acids and include modified sequences, such as glycoproteins.

Accordingly, the polypeptides of the invention encompass naturally occurring proteins, as well as those which are recombinantly or synthetically synthesized. Polypeptide or protein fragments are also encompassed by the invention. Fragments can have the same or substantially the same amino acid sequence as the naturally occurring protein. In general, polypeptides of the invention include peptides, or full-length proteins, that contain substitutions, deletions, or insertions into the protein backbone, that would still have an approximately 70%, 80%, 85%, 90%, 95% or even 100% sequence identity to the original protein over the corresponding portion.

A conservative variation denotes the replacement of an amino acid residue by a biologically similar residue. Examples of conservative variations include the substitution of one hydrophobic residue, such as isoleucine, valine, leucine or methionine, for another hydrophobic residue, or the substitution of one polar residue for another polar residue, such as the substitution of arginine for lysine, glutamic for aspartic acids, or glutamine for asparagine. Other illustrative examples of conservative substitutions include the following: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine;

isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine, glutamine, or glutamate; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

5           The present invention encompasses sequence alterations that increase stability or enhance solubility.

### **Nucleic Acids, Cloning and Expression Systems**

10           The present disclosure further provides isolated nucleic acids encoding the disclosed polyketide synthases, and active fragments thereof. The nucleic acids may comprise DNA or RNA and may be wholly or partially synthetic or recombinant. Reference to a nucleotide sequence as set out herein encompasses a DNA molecule with the specified sequence, and encompasses a RNA molecule with the specified sequence in which U is substituted for T, unless context requires otherwise.

15           The present disclosure also provides constructs in the form of plasmids, vectors, phagemids, transcription or expression cassettes that comprise at least one nucleic acid encoding a polyketide synthase or a fragment thereof, disclosed herein.

          The disclosure further provides a host cell that comprises one or more constructs as above.

20           Also provided are methods of making the encoded products. The method comprises expressing the encoded product from the encoding nucleic acid. Expression may be achieved by culturing under appropriate conditions recombinant host cells containing the nucleic acid. Following production by expression a polyketide synthase may be isolated and/or purified using any suitable technique, then used as appropriate.

25           Polyketide synthases, active fragments, and encoding nucleic acid molecules and vectors may be isolated and/or purified from their natural environment, in substantially pure or homogeneous form, or, in the case of nucleic acid, free or substantially free of nucleic acid or genes of origin other than the sequence encoding a polypeptide with the required function.

30           Systems for cloning and expression of a polypeptide in a variety of different host cells are well known in the art. For cells suitable for producing antibodies, see *Gene Expression Systems*, Academic Press, eds. Fernandez *et al.*, 1999. Briefly, suitable host cells include, but are not

limited to yeast, plant, algae, bacterial, mammalian, and insect cells. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, NS0 mouse myeloma cells, and many others. A common bacterial host is *E. coli*. Any protein expression system compatible with the invention  
5 may be used to produce the disclosed antibodies. Suitable expression systems include transgenic animals described in *Gene Expression Systems*, Academic Press, eds. Fernandez *et al.*, 1999.

Suitable vectors can be chosen or constructed, so that they contain appropriate regulatory sequences, including promoter sequences, terminator sequences, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids  
10 or viral, *e.g.*, phage, or phagemid, as appropriate. For further details see, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989. Many known techniques and protocols for manipulation of nucleic acid, for example, in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in*  
15 *Molecular Biology*, 2nd Edition, eds. Ausubel *et al.*, John Wiley & Sons, 1992.

A further aspect of the disclosure provides a host cell comprising a nucleic acid as disclosed here. A still further aspect provides a method comprising introducing such nucleic acid into a host cell. The introduction may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran,  
20 electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, *e.g.*, vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. The introduction of the nucleic acid into the cells may be followed by causing or allowing expression from the nucleic acid, *e.g.*, by culturing host cells under conditions for expression of  
25 the gene.

A wide variety of host cells are available for expressing polyketide synthase mutants of the present invention. Such host cells include, for example, yeast, plant, algae, bacterial, mammalian, and insect cells.

Once a synthase of the present invention is expressed, the protein obtained therefrom can  
30 be purified so that structural analysis, modeling, and/or biochemical analysis can be performed, as exemplified herein. The nature of the protein obtained can be dependent on the expression

system used. For example, genes, when expressed in mammalian or other eukaryotic cells, may contain latent signal sequences that may result in glycosylation, phosphorylation, or other post-translational modifications, which may or may not alter function. Therefore, a preferred embodiment of the present invention is the expression of mutant synthase genes in *E. coli* cells.

5 Once said proteins are expressed, they can be easily purified using techniques common to the person having ordinary skill in the art of protein biochemistry, such as, for example, techniques described in Colligan et al, (1997) *Current Protocols in Protein Science*, Chanda, V. B., d., John Wiley & Sons, Inc., which is incorporated by reference herein. Such techniques often include the use of cation-exchange or anion-exchange chromatography, gel filtration-size exclusion  
10 chromatography, and the like. Another technique that may be commonly used is affinity chromatography. Affinity chromatography can include the use of antibodies, substrate analogs, or histidine residues (His-tag technology).

### **Expression of 2-Pyrone Synthase in Plants**

15 Plants (including isolated plant cells) may be modified to express a 2-pyrone synthase variant (e.g., a variant comprising one or more mutations described herein). A transgenic plant, or population of such plants, expressing a 2-pyrone synthase variant transgene (e.g., a 2-pyrone synthase variant polypeptide or 2-pyrone synthase variant nucleic acid molecule) are useful in the methods of the invention. If desired, a plant expressing a 2-pyrone synthase variant of the  
20 invention includes one or more other transgenes.

In one embodiment, a 2-pyrone synthase variant nucleic acid molecule is expressed by a stably-transfected plant cell line, a transiently-transfected plant cell line, or by a transgenic plant. A number of vectors suitable for stable or extrachromosomal transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are described in  
25 Pouwels et al.(supra), Weissbach and Weissbach (supra), and Gelvin et al. (supra). Methods for constructing such cell lines are described in, e. g., Weissbach and Weissbach (supra), and Gelvin et all. (supra).

Typically, plant expression vectors include (1) a cloned plant gene under the transcriptional control of 5'and 3'regulatory sequences and (2) a dominant selectable marker.

30 Such plant expression vectors may also contain, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive, pathogen-or wound-induced, environmentally-

or developmentally-regulated, or cell-or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal. In one embodiment, a 2-pyrone synthase variant polypeptide is expressed, for example, in the glandular tricone tissue of a plant.

5 If desired, a 2-Pyrone synthase variant of the invention is modified to enhance expression in a desired plant or plant cell. In one embodiment, a non-coding flanking is subjected to mutagenesis. In one embodiment, a 2-pyrone synthase variant DNA sequence of the invention may, if desired, be combined with other DNA sequences in a variety of ways. A 2-Pyrone synthase DNA sequence of the invention may be employed with all or part of the gene sequences  
10 normally associated with a 2-Pyrone synthase protein. In its component parts, a DNA sequence encoding an 2-Pyrone synthase protein is combined in a DNA construct having a transcription initiation control region capable of promoting transcription and translation in a host cell.

In general, the constructs will involve regulatory regions functional in plants which provide for modified production of a 2-Pyrone synthase protein as discussed herein. The open  
15 reading frame coding for the 2-Pyrone synthase protein or functional fragment thereof will be joined at its 5' end to a transcription initiation regulatory region. Numerous transcription initiation regions are available which provide for constitutive or inducible regulation. For applications where developmental, cell, tissue, hormonal, or environmental expression is desired, appropriate 5' upstream non-coding regions are obtained from other genes, for example, from  
20 genes regulated during meristem development, seed development, embryo development, or leaf development.

Regulatory transcript termination regions may also be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding a 2-Pyrone synthase protein or any convenient transcription termination region derived  
25 from a different gene source. In various embodiments, the transcript termination region will contain at least 1-3 kb of sequence 3' to the structural gene from which the termination region is derived. Plant expression constructs having a 2-Pyrone synthase gene as the DNA sequence of interest for expression may be employed with a wide variety of plant life, particularly plant life involved in the production of storage reserves (for example, those involving carbon and nitrogen  
30 metabolism). Such genetically-engineered plants are useful for a variety of industrial and agricultural applications. Importantly, this invention is applicable to dicotyledons and

monocotyledons, and will be readily applicable to any new or improved transformation or regeneration method.

The expression constructs include at least one promoter operably linked to at least one 2-Pyrone synthase gene. An example of a useful plant promoter according to the invention is a  
5 caulimovirus promoter, for example, a cauliflower mosaic virus (CaMV) promoter. These promoters confer high levels of expression in most plant tissues, and the activity of these promoters is not dependent on virally encoded proteins. CaMV is a source for both the 35S and 19S promoters.

Examples of plant expression constructs using these promoters are found in Fraley et al. ,  
10 U. S. Pat. No. 5,352, 605. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odell et al., Nature 313:810, 1985). The CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., Plant Cell 2 : 591, 1990 ; Terada and Shimamoto, Mol. Gen. t Genet. 220: 389, 1990). Moreover, activity of this promoter can be  
15 further increased (i. e. , between 2-10 fold) by duplication of the CaMV 35S promoter (see e.g., Kay et al., Science 236: 1299, 1987 ; Ow et al., Proc. Natl. Acad. Sci., U. S. A. 84:4870, 1987; and Fang et al., Plant Cell 1 : 141, 1989, and McPherson and Kay, U. S. Pat. No. 5,378, 142). Other useful plant promoters include, without limitation, the nopaline synthase (NOS) promoter (An et al., Plant Physiol. 88: 547, 1988 and Rodgers and Fraley, U. S. Pat. No. 5,034, 322), the octopine synthase promoter (Fromm et al., Plant Cell 1 : 977, 1989), figwort mosaic virus (FMV) promoter  
20 (Rodgers, U. S. Pat. No. 5,378, 619), and the rice actin promoter (Wu and McElroy, W091/09948). Exemplary monocot promoters include, without limitation, commelina yellow mottle virus promoter, sugar cane badna virus promoter, rice tungro bacilliform virus promoter, maize streak virus element, and wheat dwarf virus promoter.

For certain applications, it may be desirable to produce the 2-Pyrone synthase gene  
25 product in an appropriate tissue, such as glandular trichome tissue, at an appropriate level, or at an appropriate developmental time. For this purpose, there are an assortment of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, shown to be regulated in response to inducible signals such as the environment, hormones, and/or  
developmental cues. These include, without limitation, gene promoters that are responsible for  
30 heat-regulated gene expression (see, e.g., Callis et al., Plant Physiol. 88: 965, 1988 ; Takahashi and Komeda, Mol. Gen. Genet. 219: 365, 1989 ; and Takahashi et al. Plant J. 2: 751, 1992), light-

regulated gene expression (*e.g.*, the pea *rbcS*-3A described by Kuhlemeier et al., Plant Cell 1 : 471,1989 ; the maize *rbcS* promoter described by Schaffner and Sheen, Plant Cell 3 : 997,1991 ; the chlorophyll a/b-binding protein gene found in pea described by Simpson et al., EMBO J. 4: 2723,1985 ; the Arabidopsis promoter; or the rice *rbS* promoter), hormone-regulated gene expression (for example, the abscisic acid (ABA) responsive sequences from the *Em* gene of wheat described by Marcotte et al., Plant Cell 1 : 969,1989 ; the ABA-inducible HVA1 and HVA22, and *rd29A* promoters described for barley and Arabidopsis by Straub et al., Plant Cell 6 : 617,1994 and Shen et al., Plant Cell 7 : 295,1995, organ-specific gene expression (for example, of the tuber-specific storage protein gene described by Roshal et al., EMBO J. 6:1155, 1987; the 23-kDa zein gene from maize described by Schernthaner et al. , EMBO J. 7: 1249,1988 ; or the French bean B- phaseolin gene described by Bustos et al., Plant Cell 1 : 839,1989.

Plant expression vectors may also optionally include RNA processing signals, *e.g.*, introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., Genes and Dev. 1: 1183,1987). The location of the RNA splice sequences can dramatically influence the level of transgene expression in plants. In view of this fact, an intron may be positioned upstream or downstream of an 2-Pyrone synthase polypeptide-encoding sequence in the transgene to modulate levels of gene expression.

In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally present in the 3' regions of plant genes (Thornburg et al., Proc.Natl. Acad. Sci. U.S.A. 84:744, 1987; An et al., Plant Cell 1 : 115, 1989). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the PI-11 terminator region of potato. In addition, other commonly used terminators are derived from the octopine or nopaline synthase signals. The plant expression vector also typically contains a dominant selectable marker gene used to identify those cells that have become transformed. Useful selectable genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin, or spectinomycin. Genes required for photosynthesis may also be used as selectable markers in photosynthetic- deficient strains. Finally, genes encoding herbicide resistance may be used as selectable markers ; useful herbicide resistance genes include the *bar* gene encoding the enzyme

phosphinothricin acetyltransferase and conferring resistance to the broad spectrum herbicide Basta (Frankfurt, Germany).

Efficient use of selectable markers is facilitated by a determination of the susceptibility of a plant cell to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100  $\mu\text{g/mL}$  (kanamycin), 20-50  $\mu\text{g/mL}$  (hygromycin), or 5-10  $\mu\text{g/mL}$  (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., by Vasil et al., supra. In addition, if desired, the plant expression construct may contain a modified or fully-synthetic structural 2-Pyrone synthase coding sequence that has been changed to enhance the performance of the gene in plants. Methods for constructing such a modified or synthetic gene are described in Fischhoff and Perlak, U. S. Pat. No. 5,500, 365. It should be readily apparent to one skilled in the art of molecular biology, especially in the field of plant molecular biology, that the level of gene expression is dependent, not only on the combination of promoters, RNA processing signals, and terminator elements, but also on how these elements are used to increase the levels of selectable marker gene expression.

### Plant Transformation

Upon construction of the plant expression vector, several standard methods are available for introduction of the vector into a plant host, thereby generating a transgenic plant. These methods include (1) *Agrobacterium*- mediated transformation (*A. tumefaciens* or *A. rhizogenes*) (see, e. g., Lichtenstein and Fuller In: Genetic Engineering, vol 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C. P. , and Draper, J, . In: DNA Cloning, Vol II, D. M. Glover, ed, Oxford, IRI Press, 1985)), (2) the particle delivery system (see, e.g., Gordon-Kamm et al., Plant Cell 2 : 603 (1990); or BioRad Technical Bulletin 1687, supra), (3) microinjection protocols (see, e. g., Green et al., supra), (4) polyethylene glycol (PEG) procedures (see, e. g., Draper et al., Plant Cell Physiol. 23: 451, 1982 ; or e.g., Zhang and Wu, Theor. Appl. Genet. 76: 835, 1988), (5) liposome-mediated DNA uptake (see, e. g., Freeman al., Plant Cell Physiol. 25: 1353, 1984), (6) electroporation protocols (see, e.g., Gelvin et al., supra ; Dekeyser et al., supra ; Fromm et al., Nature 319: 791, 1986 ; Sheen Plant Cell 2 : 1027, 1990 ; or Jang and Sheen Plant Cell 6 : 1665, 1994), and (7) the vortexing method (see, e. g., Kindle supra).

The method of transformation is not critical to the invention. Any method which provides for efficient transformation may be employed. As newer methods are available to transform crops or other host cells, they may be directly applied. Suitable plants for use in the practice of the invention include, but are not limited to, artemesia, sugar cane, wheat, rice, maize, sugar beet, potato, barley, manioc, sweet potato, soybean, sorghum, cassava, banana, grape, oats, solanaceous plants (e.g., tomato), millet, coconut, orange, rye, cabbage, apple, watermelon, canola, cotton, carrot, garlic, onion, pepper, strawberry, yam, peanut, onion, bean, pea, mango, citrus plants, walnuts, and sunflower.

The following is an example outlining one particular technique, an *Agrobacterium*-mediated plant transformation. By this technique, the general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, cloning and DNA modification steps are carried out in *E. coli*, and the plasmid containing the gene construct of interest is transferred by conjugation or electroporation into *Agrobacterium*. Second, the resulting *Agrobacterium* strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains an origin of replication that allows it to replicate in *Agrobacterium* and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to *Agrobacterium* for subsequent introduction into plants. Resistance genes can be carried on the vector, one for selection in bacteria, for example, streptomycin, and another that will function in plants, for example, a gene encoding kanamycin resistance or herbicide resistance. Also present on the vector are restriction endonuclease sites for the addition of one or more transgenes and directional T-DNA border sequences which, when recognized by the transfer functions of *Agrobacterium*, delimit the DNA region that will be transferred to the plant.

In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad) used for the shooting, a gunpowder charge (22 caliber Power Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The latter is fired at an acrylic stopping plate that has a hole through it that is too small for the macroprojectile to pass through. As a result, the plastic macroprojectile smashes against the stopping plate, and the tungsten microprojectiles continue toward their target through

the hole in the plate. For the instant invention the target can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast. In general, transfer and expression of transgenes in plant cells are now routine for one skilled in the art, and have become major tools to carry out gene expression studies in plants and to produce improved plant varieties of agricultural or commercial interest.

### Transgenic Plant Regeneration

Plant cells transformed with a plant expression vector can be regenerated, for example, from single cells, callus tissue, or leaf discs according to standard plant tissue culture techniques.

It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil supra ; Green et al., supra ; Weissbach and Weissbach, supra ; and Gelvin et al., supra. In one particular example, a cloned 2-Pyrone synthase polypeptide expression construct under the control of the 35SCaMV promoter and the nopaline synthase terminator and carrying a selectable marker (for example, kanamycin resistance) is transformed into *Agrobacterium*. Transformation of leaf discs, with vector-containing *Agrobacterium* is carried out as described by Horsch et al. (Science 227: 1229,1985). Putative transformants are selected after a few weeks (for example, 3 to 5 weeks) on plant tissue culture media containing kanamycin (e. g. 100Lg/nlL). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin resistant plants are then selected for greenhouse growth. If desired, seeds from self-fertilized transgenic plants can then be sowed in a soil-less medium and grown in a greenhouse. Kanamycin-resistant progeny are selected by sowing surfaced sterilized seeds on hormone-free kanamycin-containing media.

Analysis for the integration of the transgene is accomplished by standard techniques (see, for example, Ausubel et al. supra ; Gelvin et al. supra). Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA by standard immunoblot and DNA detection techniques. Each positive transgenic plant and its transgenic progeny are unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic DNA is in most cases random, and the site of integration can profoundly affect the levels and the tissue and developmental patterns of transgene expression. Consequently, a number of transgenic lines are usually screened for each

transgene to identify and select plants with the most appropriate expression profiles.

### **Transgenic plant lines are evaluated for levels of transgene expression.**

Pyrene synthase variant expression at the RNA level is determined initially to identify  
5 and quantitate expression-positive plants. Standard techniques for RNA analysis are employed  
for transgenic plants expressing 2-Pyrene synthase nucleic acids. Such techniques include PCR  
amplification assays using oligonucleotide primers designed to amplify only transgene RNA  
templates and solution hybridization assays using transgene- specific probes (see, e.g., Ausubel  
et al., supra). Those RNA-positive plants that encode a 2-Pyrene synthase protein are then  
10 analyzed for protein expression by Western immunoblot analysis using 2-Pyrene synthase  
specific antibodies (see, e.g., Ausubel et al., supra). In addition, in situ hybridization and  
immunocytochemistry according to standard protocols can be done using transgene-specific  
nucleotide probes and antibodies, respectively, to localize sites of expression within transgenic  
tissue.

### **Yeast Cells and Expression**

In one embodiment, 2-pyrene synthase variants having increased activity are expressed in  
yeast cells. Exemplary species include *Kluyveri lactis*, *Schizosaccharomyces pombe*, *Ustilago*  
*maydis*, and *Saccharomyces cerevisiae*. Other yeast which can be used in practicing the  
20 invention include, but are not limited to, *Neurospora crassa*, *Aspergillus niger*, *Aspergillus*  
*nidulans*, *Pichia pastoris*, *Candida tropicalis*, and *Hansenula polymorpha*. The term "yeast", as  
used herein, includes not only yeast in a strictly taxonomic sense, i.e., unicellular organisms, but  
also yeast-like multicellular fungi or filamentous fungi.

If desired, the yeast strain is subjected to genetic selection. Such methods are known in  
25 the art. In one embodiment, nutritional selection is used. For example, yeast strains that are  
auxotrophic for histidine (HIS 3) are known, see Struhl and Hill, (1987) Mol. Cell. Biol., 7:104;  
Fasullo and Davis, Mol. Cell. Biol., (1988) 8:4370), and the HIS3 (imidazoleglycerol phosphate  
dehydratase) gene has been used as a selective marker in yeast. See Sikorski and Heiter, (1989)  
Genetics, 122:19; Struhl, et al., P.N.A.S. (1979) 76:1035; and, for FUS1-HIS3 fusions, see  
30 Stevenson, et al., (1992) Genes Dev., 6:1293.

In certain embodiments, the host yeast cell can be modified in other ways. For example, it may be desirable to inactivate, such as by mutation or deletion, an endogenous yeast polypeptide.

In general, it will be desirable that an expression vector be capable of replication in a host cell. Heterologous DNA may be integrated into the host genome, and thereafter replicated as a  
5 part of the chromosomal DNA, or it may be DNA which replicates autonomously, as in the case of a plasmid. In the latter case, the vector will include an origin of replication which is functional in the host. In the case of an integrating vector, the vector may include sequences which facilitate integration, e.g., sequences homologous to host sequences, or encoding integrases.

Appropriate cloning and expression vectors for use with bacterial, fungal, plant, insect,  
10 yeast, and mammalian cellular hosts are known in the art, and are described in, for example, Powels et al. (Cloning Vectors: A Laboratory Manual, Elsevier, N.Y., 1985). A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach et al. (1983) in  
15 Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due to the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid.

In addition, drug resistance markers such as ampicillin can be used. Suitable promoters  
20 for function in yeast include the promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., J. Biol. Chem. 255, 2073 (1980) or other glycolytic enzymes (Hess et al., J. Adv. Enzyme Req. 7, 149 (1968); and Holland et al. Biochemistry 17, 4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phospho-fructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate  
25 kinase, triosephosphate isomerase, phospho-glucose isomerase, and glucokinase. Suitable vectors and promoters for use in yeast expression are further described in R. Hitzeman et al., EPO Pub. No. 73,657.

## Mutagenesis and Expression of Biomolecules of Interest

Various types of cloning and mutagenesis methods can be used with the methods herein, *e.g.*, to produce and/or isolate polyketide synthases and/or to further modify/mutate the polypeptides (*e.g.*, polyketide synthases) provided herein.

5 In some embodiments, isolated nucleic acids, polypeptides and/or viruses can be further mutated. Mutagenesis methods include but are not limited to site-directed, random point mutagenesis, homologous recombination (DNA shuffling), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA and the like. Additional suitable mutagenesis methods  
10 include point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, *e.g.*, involving chimeric constructs, is also included in the methods herein. In some embodiments, mutagenesis can be guided by known information of the naturally occurring molecule or altered or mutated naturally occurring  
15 molecule, *e.g.*, sequence, sequence comparisons, physical properties, crystal structure or the like.

Oligonucleotides for use in mutagenesis are typically synthesized chemically according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers, Tetrahedron Letts 22(20):1859-1862, (1981) using an automated synthesizer, as described in Needham-VanDevanter *et al.*, Nucleic Acids Res, 12:6159-6168 (1984). In addition, essentially  
20 any nucleic acid can be custom or standard ordered from any of a variety of commercial sources. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources.

Also provided herein are host cells and organisms comprising a polyketide synthase or other polypeptide and/or nucleic acid provided herein or such polyketide synthase or other sequences within various vectors such as viruses, plasmids in plasmid rescue systems, and the  
25 like. Host cells can be transformed, transduced or transfected with the vectors provided herein, which can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors can be introduced into cells and/or microorganisms by standard methods including electroporation (see, From *et al.*, Proc Natl Acad Sci USA 82, 5824 (1985)),  
30 infection by viral vectors, high velocity ballistic penetration by small particles with the nucleic

acid either within the matrix of small beads or particles, or on the surface (Klein *et al.*, Nature 327, 70-73 (1987)).

Several well-known methods of introducing target nucleic acids into bacterial cells are available, any of which can be used with the methods herein. These include, for example, fusion  
5 of the recipient cells with bacterial protoplasts containing the DNA, electroporation, projectile bombardment, and infection with viral vectors, and the like. Bacterial cells can be used to amplify the number of plasmids containing DNA constructs. The bacteria are grown to log phase and the plasmids within the bacteria can be isolated by a variety of methods known in the art. In addition, a plethora of kits are commercially available for the purification of plasmids  
10 from bacteria, (see, *e.g.*, EasyPrep.TM., FlexiPrep.TM., both from Pharmacia Biotech; StrataClean.TM., from Stratagene; and, QIAprep.TM. from Qiagen). The isolated and purified plasmids are then further manipulated to produce other plasmids, used to transfect cells or incorporated into related vectors to infect organisms. Typical vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful  
15 for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both, (*e.g.*, shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or both. See, Gilman &  
20 Smith, Gene 8:81 (1979); Roberts, *et al.*, Nature, 328:731 (1987); Schneider, B., *et al.*, Protein Expr Purif 6435:10 (1995); Ausubel, Sambrook, Berger (all supra). A catalogue of bacteria and bacteriophages useful for cloning is provided, *e.g.*, by the ATCC, *e.g.*, The ATCC Catalogue of Bacteria and Bacteriophage (1992) Gherna *et al.* (eds.) published by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying  
25 theoretical considerations are known in the art.

According to the present invention, nucleic acid sequences encoding a mutated polyketide synthase can be produced by the methods described herein, or any alternative methods available to the skilled artisan. In designing the nucleic acid sequence of interest, it  
30 maybe desirable to reengineer said gene for improved expression in a particular expression system. For example, it has been shown that many bacterially derived genes do not express well in yeast systems, in some cases, yeast-derived genes do not express well in bacteria. This

phenomenon may be due to the non-optimal G+C content and/or A+T content of said gene relative to the expression system being used. Thus, one goal in the design of genes is to generate nucleic acid sequences that have a G+C content that affords mRNA stability and translation accuracy for a particular expression system.

5           Due to the redundancy of the genetic code (*i.e.*, many amino acids are specified by more than one codon), evolution of the genomes of different organisms or classes of organisms has resulted in differential usage of redundant codons. This “codon bias” is reflected in the mean base composition of protein coding regions. For example, organisms with relatively low G+C contents utilize codons having A or T in the third position of redundant codons, whereas those  
10           having higher G+C contents utilize codons having G or C in the third position.

          Therefore, in reengineering genes for expression, one may wish to determine the codon bias of the organism in which the gene is to be expressed. Looking at the usage of the codons as determined for genes of a particular organism deposited in GenBank can provide this information. After determining the bias thereof, the new gene sequence can be analyzed for  
15           restriction enzyme sites as well as other sites that could affect transcription such as exon intron junctions, polyA addition signals, or RNA polymerase termination signals.

          Once expressed, mutants of the present invention may be purified for characterization by any of several different properties. In some embodiments, such mutants may have altered active site surface charges of one or more charge units. In addition, said mutants may have altered  
20           substrate specificity or product capability relative to a non-mutated polyketide synthase. In other embodiments, such mutants may have altered stability.

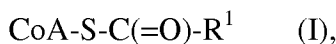
### **Compounds and Methods of Making Same**

          The present invention includes methods of preparing compounds using the synthases and  
25           substrates contemplated within the invention. The present invention further includes compounds that are prepared using the synthases and substrates contemplated within the invention. The invention further includes derivatives of compounds that are prepared using the synthases and substrates contemplated within the invention, such as, but not limited to, partially or fully alkylated, or partially or fully acylated derivatives thereof.

30           The preparation of a compound contemplated within the invention includes the step of contacting a substrate contemplated within the invention with a synthase contemplated within the

invention, wherein the synthase catalyzes reaction of the substrate. In certain embodiments, the substrate comprises acetyl-CoA. In other embodiments, the substrate comprises acetoacetyl-CoA, malonyl-CoA, or coumaryl-CoA. In yet other embodiments, the substrate comprises a compound of formula (I), or a salt, solvate or derivative thereof:

5



wherein in (I)  $\text{R}^1$  is selected from the group consisting of  $\text{C}_1\text{-C}_{20}$  alkyl,  $\text{C}_1\text{-C}_{20}$  cycloalkyl, aryl-  
( $\text{C}_1\text{-C}_{20}$  alkyl), carboxy-( $\text{C}_1\text{-C}_{20}$  alkyl), aryl, heterocyclyl, heteroaryl, aryl-CH=CH-,  
10 heterocyclyl-CH=CH-, and heteroaryl-CH=CH-,

wherein each alkyl, cycloalkyl, aryl, heterocyclyl or heteroaryl group is independently optionally substituted with at least one substituent selected from the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro, monoalkylamino, dialkylamino, carboxy,  $\text{C(=O)O(C}_1\text{-C}_6\text{)alkyl}$ , trifluoromethyl, cyano,  $\text{-C(=O)NH}_2$ ,  $\text{-C(=O)NH(C}_1\text{-C}_6\text{)alkyl}$ ,  
15  $\text{-C(=O)N((C}_1\text{-C}_6\text{)alkyl)}_2$ ,  $\text{-SO}_2\text{NH}_2$ , and  $\text{-C(=NH)NH}_2$ .

In certain embodiments,  $\text{R}^1$  is selected from the group consisting of  $\text{C}_1\text{-C}_{20}$  alkyl, carboxy-( $\text{C}_1\text{-C}_{20}$  alkyl), aryl, heteroaryl, aryl-CH=CH- and heteroaryl-CH=CH-. In certain embodiments, the alkyl, cycloalkyl, aryl or heteroaryl group is optionally substituted with at least one substituent selected from the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro,  
20 monoalkylamino, dialkylamino, carboxy,  $\text{C(=O)O(C}_1\text{-C}_6\text{)alkyl}$ , trifluoromethyl, cyano,  $\text{-C(=O)NH}_2$ ,  $\text{-C(=O)NH(C}_1\text{-C}_6\text{)alkyl}$ ,  $\text{-C(=O)N((C}_1\text{-C}_6\text{)alkyl)}_2$ ,  $\text{-SO}_2\text{NH}_2$ , and  $\text{-C(=NH)NH}_2$ . In certain embodiments, the aryl group is phenyl. In certain embodiments, the heteroaryl group is pyridyl.

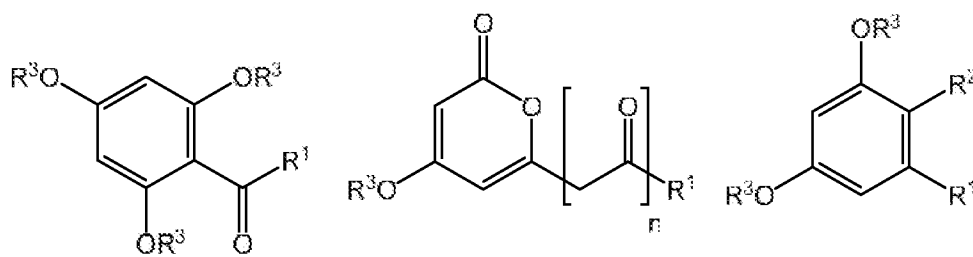
The substrate comprising the compound of formula (I) may be prepared using methods  
25 known to those skilled in the art, such as but not limited to acylation of CoA using the appropriate acylating agent under conditions known to those skilled in the art.

Upon contacting the substrate and the synthase contemplated within the invention, the substrate undergoes a chemical reaction that is catalyzed by the synthase. In certain embodiments, the product of the synthase-catalyzed substrate reaction is a compound  
30 contemplated within the invention. In other embodiments, the product of the synthase-catalyzed substrate reaction is further derivatized, and the resulting derivatization product is a compound

contemplated within the invention. The derivatization of any compound may be performed using chemical methods (i.e., reacting the product of the synthase-catalyzed substrate reaction with one or more chemical reagents, either simultaneously or in sequence), enzymatic methods (i.e., contacting the product of the synthase-catalyzed substrate reaction with one or more enzymes, such as but not limited to other synthases, chalcones and the like, either simultaneously or in sequence), and/or a combination of chemical methods and biochemical methods.

The product of the synthase-catalyzed substrate reaction and/or the resulting product from the derivatization of the product of the synthase-catalyzed substrate reaction may be isolated using methods known to those skilled in the art, such as but not limited to chromatography, liquid-liquid or liquid-solid extraction, precipitation, evaporation and/or sublimation.

The invention contemplates compounds generated using the naturally occurring or mutant synthases of the invention. In certain embodiments, the compounds contemplated within the invention comprise a cyclic compound, or a salt, solvate or derivative thereof, selected from the group consisting of:



wherein:

$R^1$  is selected from the group consisting of  $C_1$ - $C_{20}$  alkyl,  $C_1$ - $C_{20}$  cycloalkyl, aryl- $(C_1$ - $C_{20}$  alkyl), carboxy- $(C_1$ - $C_{20}$  alkyl), aryl, heterocyclyl, heteroaryl, aryl-CH=CH-, heterocyclyl-CH=CH- and heteroaryl-CH=CH-,

$R^2$  is selected from the group consisting of H, carboxy,  $C(=O)O(C_1$ - $C_6$ )alkyl, cyano,  $-C(=O)NH_2$ ,  $-C(=O)NH(C_1$ - $C_6$ )alkyl, and  $-C(=O)N((C_1$ - $C_6$ )alkyl) $_2$ ;

each occurrence of  $R^3$  is independently selected from the group consisting of H,  $C_1$ - $C_{20}$  alkyl, aryl- $(C_1$ - $C_{20}$  alkyl),  $C_1$ - $C_{20}$  acyl, aroyl, and heteroaroyl; and,

$n$  is 0, 1, 2 or 3;

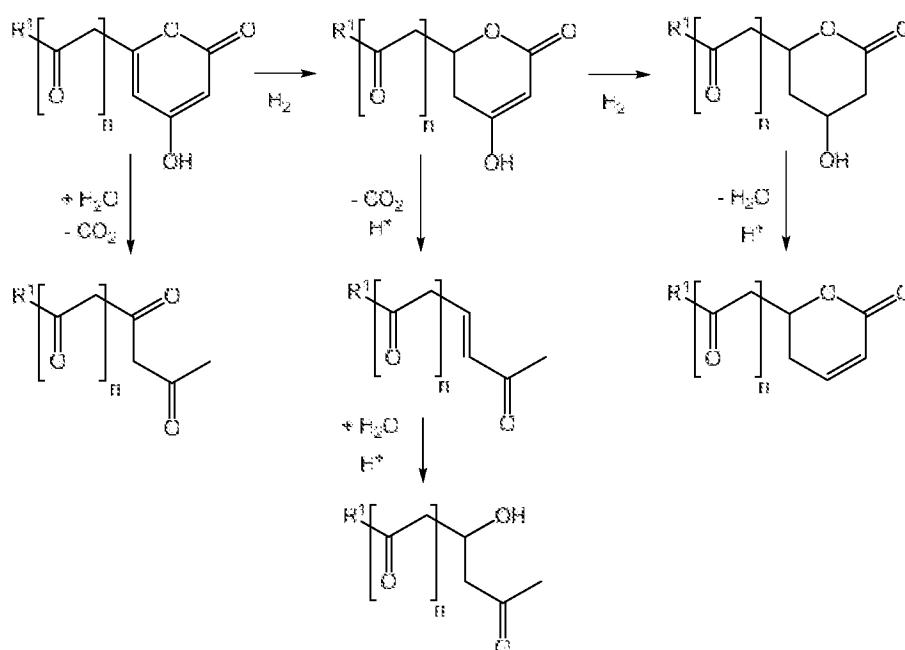
wherein each alkyl, cycloalkyl, aryl, heterocyclyl or heteroaryl group is independently optionally substituted with at least one substituent selected from

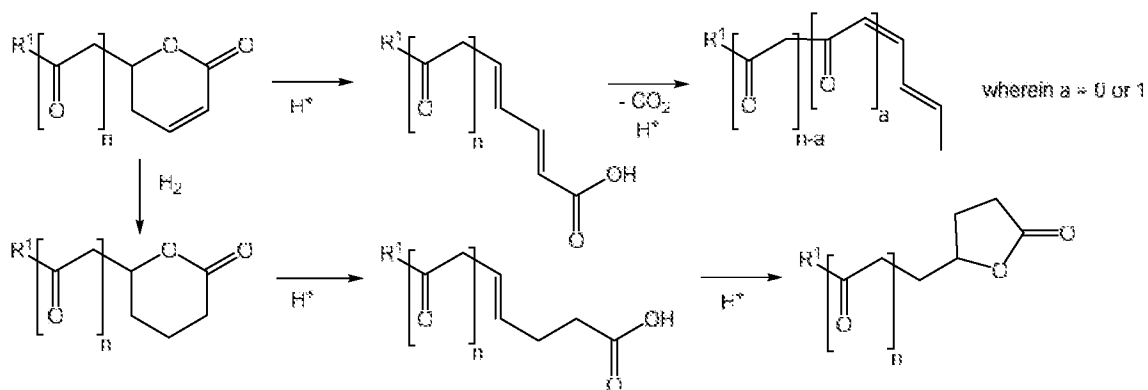
the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro, monoalkylamino, dialkylamino, carboxy,  $C(=O)O(C_1-C_6)alkyl$ , trifluoromethyl, cyano,  $-C(=O)NH_2$ ,  $-C(=O)NH(C_1-C_6)alkyl$ ,  $-C(=O)N((C_1-C_6)alkyl)_2$ ,  $-SO_2NH_2$ , and  $-C(=NH)NH_2$ .

5 In certain embodiments, the compound obtained by contacting a compound of formula (I) with a synthase contemplated within the invention may be further derivatized by, such as but not limited to, partial or total alkylation, or partial or total acylation. Any chemical products of such derivatization reactions are also compounds contemplated within the invention.

10 In certain embodiments, the compound obtained by contacting acetyl-CoA or any known acylated CoA derivative with a synthase contemplated within the invention may be further derivatized by, such as but not limited to, partial or total alkylation, or partial or total acylation; any products of such derivatization reactions are also compounds contemplated within the invention.

15 In certain embodiments, the compounds of the invention may be further derivatized using known procedures known in the art, and any products obtained by such derivatization reactions are also compounds contemplated within the present invention. In a non-limiting example, all compounds illustrated in the following scheme are compounds contemplated within the present invention.





wherein  $R^1$  is selected from the group consisting of  $C_1$ - $C_{20}$  alkyl,  $C_1$ - $C_{20}$  cycloalkyl, aryl- $(C_1$ - $C_{20}$  alkyl), carboxy- $(C_1$ - $C_{20}$  alkyl), aryl, heterocyclyl, heteroaryl, aryl- $CH=CH$ -, heterocyclyl- $CH=CH$ - and heteroaryl- $CH=CH$ -, wherein each alkyl, cycloalkyl, aryl, heterocyclyl or heteroaryl group is independently optionally substituted with at least one substituent selected from the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro, monoalkylamino, dialkylamino, carboxy,  $C(=O)O(C_1-C_6)alkyl$ , trifluoromethyl, cyano,  $-C(=O)NH_2$ ,  $-C(=O)NH(C_1-C_6)alkyl$ ,  $-C(=O)N((C_1-C_6)alkyl)_2$ ,  $-SO_2NH_2$ , and  $-C(=NH)NH_2$ .

In certain embodiments, at least one functional group of the compound of the invention is protected with an appropriate protective group before the compound of the invention is further derivatized, as to avoid premature reaction of the at least one functional group being protected. The protective groups that are useful within the invention are well-known to those skilled in the art, and are determined in view of the functional group reactivity and the derivatization conditions used. In certain embodiments, the functional groups of the compounds of the invention are not protected with a protective group before the compound of the invention is further derivatized. In the event that derivatization has the effect of modifying any functional group(s) of the compounds of the invention, the resulting modified products are also compounds of the present invention.

The compounds of the present invention may be formulated as pharmaceutically acceptable compositions, wherein the compositions further comprise at least one pharmaceutically acceptable carrier.

## Structural and Computational Methods

The three-dimensional structures of polyketide synthase proteins have been determined. In the present study, new structures of 2-pyrone synthase were obtained. The invention encompasses specific three-dimensional space coordinates of at least fourteen  $\alpha$ -carbon atoms defined for the active site. R-groups attached to said  $\alpha$ -carbons are defined such that mutants can be made by changing at least one R-group found in the synthase active site. Such mutants may have unique and useful properties within the present invention.

One aspect of the invention resides in obtaining crystals of the polyketide synthase polypeptide, of sufficient quality to determine the three dimensional (tertiary) structure of the protein by X-ray diffraction methods. The knowledge obtained concerning the three-dimensional structure of chalcone synthase can be used in the determination of the three-dimensional structure of other synthase polypeptides in the polyketide synthesis pathway. The structural coordinates of a polyketide synthase can be used to develop new polyketide synthesis enzymes or synthase inhibitors using various computer modeling protocols. Based on the structural coordinates of a polyketide synthase polypeptide (*e.g.*, the three-dimensional protein structure), novel polyketide synthases can be engineered. In addition, small molecules that mimic or interact with a synthase molecule functional domain can be designed and synthesized to modulate chalcone synthase, pyrone synthase, and other polyketide synthase biological functions, as well as the biological functions of other polyketide synthases.

Accordingly, in one embodiment, the invention provides a method of “rational” enzyme or drug design. Another approach to “rational” enzyme or drug design is based on a lead compound discovered using high throughput screening. This lead compound can be further modified in order to optimize its binding and/or developability properties, based on a crystal structure of the binding regions of the molecule in question. Accordingly, another aspect of the invention is to provide related protein sequences or materials that are a starting material in the rational design of new synthases or drugs, which can then lead to the synthesis of new polyketides or the rational modification of the polyketide synthesis pathway.

The present invention relates to crystallized polyketide synthases and mutants thereof, from which the position of specific  $\alpha$ -carbon atoms and R-groups associated therewith comprising the active site can be determined in three-dimensional space. The invention also relates to structural coordinates of said polyketide synthases, the use of said structural

coordinates to develop structural information related to polyketide synthase homologues, mutants, and the like, and to crystal forms of such synthases.

Furthermore, the invention, as disclosed herein, provides a method whereby said  $\alpha$ -carbon structural coordinates determined for atoms comprising the active site of said synthase can be used to design synthases wherein R-groups associated with active site  $\alpha$ -carbon atoms are different from the R-groups found in native 2-pyrone synthase, *e.g.*, variant synthases. In addition, the invention provides for production of variant polyketide synthases based on the structural information of synthases and for use of said mutant synthases to make a variety of polyketide-based compounds using a variety of substrates (as described in PCT Application No. PCT/US00/20674, filed July 27, 2000, incorporated by reference in its entirety herein). The present invention also provides methods of producing novel mutant polyketide synthases by comparing the crystal structures of two different polyketide synthases and identifying residues to be mutated, deleted or added.

Crystal structures are preferably obtained at a resolution of about 1.56 Å to about 3 Å for a polyketide synthase in the presence and in the absence of bound substrate or substrate analog. Coordinates for a polyketide synthase in the absence of a substrate bound in the active site have been deposited at the Protein Data Bank, accession number 1QLV. Those skilled in the art understand that a set of structure coordinates determined by X-ray crystallography is not without standard error.

In one embodiment, the active site may be altered to vary the activity of a polyketide synthase. An appropriate combination of R-groups, linked to active site  $\alpha$ -carbons, can facilitate the formation of one or more desired reaction products. The combination of R-groups selected for use in a synthase can be any combination other than the ordered arrangements of R-groups found in known native isolated polyketide synthases. Typically, R-groups found on active site  $\alpha$ -carbons are those found in naturally occurring amino acids, in some embodiments, however, R-groups other than those found in naturally occurring amino acids can be used.

The present invention permits the use of molecular design techniques to design, select, and synthesize mutant polyketide synthases that use the same substrates as the native synthases but yet produce different and/or novel polyketide compounds. Mutant proteins of the present invention, and nucleic acids encoding the same, can be designed by genetic manipulation based on structural information about polyketide synthases. For example, one or more R-groups

associated with the active site  $\alpha$ -carbon atoms of 2-pyrone synthase can be changed by altering the nucleotide sequence of the corresponding 2-pyrone synthase gene, thus making one or more variant polyketide synthases. Such genetic manipulations can be guided by structural information concerning the R-groups found in the active site  $\alpha$ -carbons when substrate is bound to the protein upon crystallization. Alternatively, mutant polyketide synthases can be prepared by standard protocols for polypeptide synthesis, as is well known in the art.

Furthermore, structural coordinates may be used to determine the structure of the crystalline form of other proteins with significant amino acid or structural homology to any functional domain of a synthase. One method that may be employed for such purpose is molecular replacement. In this method, the unknown crystal structure, whether it is another crystal form of a synthase, a synthase having a mutated active site, or the crystal of another protein with significant sequence and/or structural identity to a polyketide synthase, may be determined using a set of coordinates. This method provides sufficient structural information for the unknown crystal more efficiently than attempting to determine such information *ab initio*. In addition, this method can be used to determine whether or not a given polyketide synthase in question falls within the scope of this invention. Furthermore, polyketide synthases and variants thereof may be crystallized in the presence or absence of substrates and substrate analogs. The crystal structures of a series of complexes may then be solved by molecular replacement and compared to that of the wild-type to assist in determination of suitable replacements for R-groups within the active site, thus making synthase mutants according to the present invention.

All variants of the present inventions may be modeled using previously obtained structural information without necessarily having to crystallize and solve the structure for each and every mutant. For example, one skilled in the art may use one of several specialized computer programs to assist in the process of designing synthases having mutated active sites relative to the wild-type. Examples of such programs include: GRID (Goodford, 1985, J. Mod. Chem.:2S: 849-857), MCSS (Miranker and Karplus, 1991, Proteins: Structure, Function and Genetics, 11:29-34); AUTODOCK (Goodsell and Olsen, 1990, Proteins. Structure, Function, and Genetics, 8:195-202); and DOCK (Kuntz et al, 1982, J. Mol Biol: 161:269-288), and the like. In addition, specific computer programs are also available to evaluate specific substrate-active site interactions and the deformation energies and electrostatic interactions resulting therefrom. MODELLER is a computer program often used for homology or comparative modeling of the

three-dimensional structure of a protein (A. Sali & T.L. Blundell. J Mol.Biol 234:779-815, 1993). A sequence to be modeled is aligned with one or more known related structures and the MODELLER program is used to calculate a full-atom model, based on optimum satisfaction of spatial restraints. Such restraints can include, inter alia, homologous structures, site-directed  
5 mutagenesis, fluorescence spectroscopy results, NMR experiment results, or atom-atom potentials of mean force.

The present invention enables polyketide synthase mutants to be made and the crystal structure thereof to be solved. Moreover, by virtue of the present invention, the location of the active site and the interface of substrate therewith permit the identification of desirable R-groups  
10 for mutagenesis.

The three-dimensional coordinates of the polyketide synthase may be used to predict the activity and or substrate specificity of a protein whose primary amino acid sequence suggests that it may have polyketide synthase activity. The family of 2-pyrone synthase -related enzymes is defined, in part, by the presence of four highly conserved amino acid residues, Cys164,  
15 Phe215, His303, and Asn336.

In addition, the structural coordinates and three-dimensional models disclosed herein can be used to design or identify polyketide synthase inhibitors. Using the modeling techniques disclosed herein, potential inhibitor structures can be modeled with the polyketide synthase active site and those that appear to interact therewith can subsequently be tested in activity  
20 assays in the presence of substrate.

Methods of using crystal structure data to design binding agents or substrates are known in the art. Thus, the crystal structure data provided herein can be used in the design of new or improved inhibitors, substrates or binding agents. For example, the synthase polypeptide coordinates can be superimposed onto available coordinates of other similar enzymes to identify  
25 modifications in the active sites of the enzymes, thus allowing the novel enzyme products or the modulation of polyketide synthesis. Alternatively, the synthase polypeptide coordinates can be superimposed onto available coordinates of other similar enzymes which have substrates or inhibitors bound to them to give an approximation of the way these and related substrates or inhibitors might bind to a synthase. Alternatively, computer programs employed in the practice  
30 of rational drug design can be used to identify compounds that reproduce interaction characteristics similar to those found between a synthase polypeptide and a cocrystallized

substrate. Furthermore, detailed knowledge of the nature of binding site interactions allows for the modification of compounds to alter or improve developability parameters, such as but not limited to solubility, pharmacokinetics and the like, without affecting binding activity.

Widely available are computer programs capable of carrying out the activities necessary to design agents using the crystal structure information provided herein. Examples include, but are not limited to, the following computer programs: Catalyst Databases™ - an information retrieval program accessing chemical databases such as BioByte Master File, Derwent WDI and ACD; Catalyst HYPO™ - generates models of compounds and hypotheses to explain variations of activity with the structure of drug candidates; Ludi™ - fits molecules into the active site of a protein by identifying and matching complementary polar and hydrophobic groups; Leapfrog™ - “grows” new ligands using a genetic algorithm with parameters under the control of the user.

In addition, various general purpose machines may be used with programs written in accordance with the teachings herein, or it may be more convenient to construct more specialized apparatus to perform the operations. However, preferably the embodiment is implemented in one or more computer programs executing on programmable systems each comprising at least one processor, at least one data storage system (including volatile and non-volatile memory and or storage elements), at least one input device, and at least one output device. The program is executed on the processor to perform the functions described herein.

Each such program may be implemented in any desired computer language (including machine, assembly, high level procedural, object oriented programming languages, or the like) to communicate with a computer system. In any case, the language may be a compiled or interpreted language. The computer program will typically be stored on a storage media or device (*e.g.*, ROM, CD-ROM, or magnetic or optical media) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

Embodiments of the invention include systems (*e.g.*, internet based systems), particularly computer systems which store and manipulate coordinate and sequence information. As used herein, “a computer system” refers to the hardware components, software components, and data

storage components used to analyze coordinate and sequence information. The computer system typically includes a processor for processing, accessing and manipulating the sequence data. The processor can be any well-known type of central processing unit, such as, for example, the Pentium in from Intel Corporation, or similar processor from Sun, Motorola, Compaq, AMD or  
5 International Business Machines.

One approach enabled by this invention is to use the structure coordinates to design new enzymes capable of synthesizing novel and known polyketides. For example, polyketide synthases (PKSs) generate molecular diversity in their products by utilizing different starter molecules and by varying the final size of the polyketide chain. The structural coordinates allow  
10 the elucidation of the nature by which PKSs achieve starter molecule selectivity and control polyketide chain length. For example, by comparing the structure of chalcone synthase that yields a tetraketide product to 2-pyrone synthases that form a triketide product of the invention, it was demonstrated that 2-pyrone synthase maintains a smaller initiation/elongation cavity. Accordingly, generation of a chalcone synthase mutant with an active site sterically analogous to  
15 2-pyrone synthase results in the synthesis of a polyketide product of a different size. The subject invention allows for the strategic development and biosynthesis of more diverse polyketides and demonstrates a structural basis for control of polyketide chain length in other PKSs. In addition, the structural coordinates allow for the development of substrates or binding agents that bind to the polypeptide and alter the physical properties of the compounds in different ways, *e.g.*,  
20 solubility.

In another embodiment, one approach is to computationally screen small molecule databases for chemical entities or compounds that can bind in whole, or in part, to a polyketide synthase polypeptide or fragment thereof. In this screening, the quality of fit of such entities or compounds to the binding site may be judged either by shape complementarity or by estimated  
25 interaction energy. Meng, E. C. *et al.*, J. Comp. Chem., 13:505-524 (1992).

The design of substrates, compounds or binding agents that bind to or inhibit a polyketide synthase polypeptide according to the invention generally involves consideration of two factors. First, the substrate, compound or binding agent must be capable of physically and structurally associating with a polyketide synthase molecule. Non-covalent molecular interactions important  
30 in the association of a polyketide synthase with a substrate include hydrogen bonding, van der Waals and hydrophobic interactions, and the like.

Second, the substrate, compound or binding agent must be able to assume a conformation that allows it to associate with a polyketide synthase molecule. Although certain portions of the substrate, compound or binding agent will not directly participate in this association, those portions may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on potency. Such conformational requirements include the overall three-dimensional structure and orientation of the chemical entity or compound in relation to all or a portion of the binding site, *e.g.*, active site or accessory binding site of a polyketide synthase (*e.g.*, a 2-pyrone synthase polypeptide), or the spacing between functional groups of a substrate or compound comprising several chemical entities that directly interact with a polyketide synthase.

The potential binding effect of a substrate or chemical compound on a polyketide synthase, or the activity that a newly synthesized or mutated polyketide synthase might have on a known substrate, may be analyzed prior to its actual synthesis and testing by the use of computer modeling techniques. For example, if the theoretical structure of the given substrate or compound suggests insufficient interaction and association between it and a polyketide synthase, synthesis and testing of the compound may be obviated. However, if computer modeling indicates a strong interaction, the molecule may then be tested for its ability to bind to, initiate catalysis or elongation of a polyketide by a polyketide synthase. Methods of assaying for polyketide synthase activity are known in the art (as identified and discussed herein). Methods for assaying the effect of a newly created polyketide synthase or a potential substrate or binding agent can be performed in the presence of a known binding agent or polyketide synthase. For example, the effect of the potential binding agent can be assayed by measuring the ability of the potential binding agent to compete with a known substrate.

Specialized computer programs may also assist in the process of selecting fragments or chemical entities. These include:

1. GRID (Goodford, P. J., "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", *J. Med. Chem.*, 28:849-857 (1985)). GRID is available from Oxford University, Oxford, UK.
2. MCSS (Miranker, A. and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method." *Proteins: Structure. Function and Genetics*, 11:29-34 (1991)). MCSS is available from Molecular Simulations, Burlington, Mass.

3. AUTODOCK (Goodsell, D. S. and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", *Proteins: Structure, Function, and Genetics*, 8:195-202 (1990)). AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.

4. DOCK (Kuntz, I. D. *et al.*, "A Geometric Approach to Macromolecule-Ligand Interactions", *J. Mol. Biol.*, 161:269-288 (1982)). DOCK is available from University of California, San Francisco, Calif.

Once suitable substrates, chemical entities or fragments have been selected, they can be assembled into a single polypeptide, compound or binding agent (*e.g.*, an inhibitor). Assembly may be performed by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of the molecules. This would be followed by manual model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

1. CAVEAT (Bartlett, P. A. *et al.*, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules". In "Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc, 78, pp. 182-196 (1989)). CAVEAT is available from the University of California, Berkeley, Calif.

2. 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, Calif). This area is reviewed in Martin, Y. C, "3D Database Searching in Drug Design", *J Med. Chem.*, 35:2145-2154 (1992)).

3. HOOK (available from Molecular Simulations, Burlington, Mass.).

In addition to the method of building or identifying novel enzymes or a polyketide synthase substrate or binding agent in a stepwise fashion, one fragment or chemical entity at a time as described above, substrates, inhibitors or other polyketide synthase interactions may be designed as a whole or "*de novo*" using either an empty active site or optionally including some portion(s) of known substrates, binding agents or inhibitors. These methods include:

1. LUDI (Bohm, H.-L, "The Computer Program LUDI: A New Method for the *De Novo* Design of Enzyme Inhibitors", *J Comp. Aid. Molec. Design*, 6:61-78 (1992)). LUDI is available from Biosym Technologies, San Diego, Calif.

2. LEGEND (Mshibata, Y. and A. Itai, Tetrahedron, 47:8985 (1991)). LEGEND is available from Molecular Simulations, Burlington, Mass.

3. LeapFrog (available from Tripos Associates, St. Louis, Mo.).

Other molecular modeling techniques may also be employed in accordance with this invention. See, *e.g.*, Cohen, N. C. *et al.*, "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33:883-894 (1990). See also, Navia, M. A. and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2:202-210 (1992).

## 10 Kits

In various embodiments, the invention provides kits comprising a polyketide synthase variant having altered enzymatic activity and/or a host cell (*e.g.*, yeast, plant, algae, bacterial, mammalian, insect) comprising such a variant. Such kits are useful in *in vitro* and *in vivo* methods for the production of desired chemical compounds, including polyketide, pyrone and lactone products.

In some embodiments, the kit provides a sterile container comprising a composition of the invention; such containers can be boxes, ampoules, bottles, vials, tubes, bags, pouches, blister-packs, or other suitable container forms known in the art. Such containers can be made of plastic, glass, laminated paper, metal foil, or other materials suitable for holding reagents. If desired a polypeptide of the invention is provided together with instructions for pyrone production. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase

Chain Reaction”, (Mullis, 1994); “Current Protocols in Immunology” (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

5           The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the compositions of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

## EXAMPLES

### Example 1. Cloning, Expression, and Purification of 2-Pyrone synthase Enzymes

*Gerbera hybrida* 2-Pyrone synthase (Gh2PS) was subcloned into the pHIS8 expression  
5 vector, derived from pET28a(+) (Jez *et al.*, 2000a). For protein expression, Gh2PS/pHIS8-  
transformed *E. coli* BL21(DE3) cells were incubated with shaking at 37°C in Terrific-broth  
medium containing 50 µg/ml kanamycin until the OD(600nm) reached ~1.0, and then induced  
with 0.5 mM isopropyl 1-thio-β-galactopyranoside (IPTG), grown at 18°C overnight, and finally  
10 harvested by centrifugation at 8,000 x g. Cell pellets were resuspended in lysis buffer A [TNB,  
20 mM imidazole, 10% (v/v) glycerol, 1% (v/v) Tween-20, and 10 mM β-mercaptoethanol (β-  
ME); TNB is 50 mM Tris-HCl (pH 8.0), 500 mM NaCl] supplemented with 4 mM benzamidine,  
1 mM PMSF, 2 µg/ml leupeptin, and 0.5 mg/ml lysozyme. Following cell lysis by sonication  
and centrifugation at 100,000 x g, the cleared cell lysate was passed over a Ni<sup>2+</sup>-NTA column  
(Qiagen, Valencia, CA) equilibrated in lysis buffer A, which was then washed with 10 bed  
15 volumes of wash buffer (TNB, 20 mM imidazole, 10 mM β-ME). His-tagged protein was eluted  
with 10 bed volumes of elution buffer (TNB, 250 mM imidazole, 10 mM β-ME), and the N-  
terminal His-tag was cleaved by thrombin digestion during a 24-hour dialysis against digestion  
buffer (TNB, 10 mM β-ME) at 4°C. Cleaved protein was isolated by running the dialyzed  
sample over another Ni<sup>2+</sup>-NTA column equilibrated in digestion buffer to remove the His-  
20 containing peptide and any un-cleaved protein, followed by a benzamidine sepharose  
chromatography to remove thrombin. A Superdex200 or S75 gel-filtration column (Amersham)  
equilibrated in TNB with 2 mM DTT was utilized to isolate homogeneous protein, which was  
finally dialyzed against storage buffer [100 mM NaCl, 50 mM Tris-HCl (pH 8.0), 2 mM DTT],  
and prior to concentration to 10-20 mg/ml and for storage at -80°C. Site-directed mutants 2-  
25 Pyrone synthase were generated in the Gh2PS/pHIS8 expression vector, using the QuikChange  
strategy (Stratagene). The mutant proteins were expressed and purified as described above for  
the native protein.

## Example 2. Crystallization, X-ray Diffraction Data Collection, and Structure

### Determination

Crystals of wild-type and mutant Gh2PS were grown by vapor diffusion at 4°C in hanging drops, which consisted of 1 µl protein in storage buffer mixed with 1 µl crystallization reservoir [1.7 M ammonium sulfate, 7.5% (v/v) glycerol, 100 mM sodium succinate (pH 5.5), and 4 mM DTT]. Seeding was used to improve the quality and size of the crystals, which grew to 0.3 x 0.2 x 0.2 mm.

Prior to freezing in liquid nitrogen, crystals of Gh2PS were soaked for 3 minutes in cryoprotectant [crystallization reservoir supplemented with 17% (v/v) ethylene glycol]. X-ray diffraction data were collected on ADSC Q315 CCD detectors at beamlines 8.2.1 and 8.2.2 of the Advanced Light Source, Lawrence Berkeley National Laboratory.

Diffraction data were indexed, integrated, and scaled with MOSFLM (Battye *et al.*, 2011) and SCALA (Evans, 2006). Gh2PS crystallized in one of two distinct forms: space group  $P3_121$  with unit-cell dimensions  $a=b=83$  Å and  $c=242$  Å and two monomers in the asymmetric unit; or space group  $P3_221$  with unit-cell dimensions  $a=b=83$  Å and  $c=120$  Å and one monomer in the asymmetric unit.

The structures of Gh2PS proteins were determined for the  $P3_121$  crystal form from refinements initiated directly with the previously published Gh2PS homodimeric structure (PDB entry 1EE0; Jez *et al.*, 2000b); or for the  $P3_221$  crystal form through molecular-replacement (MR) analysis with the Molrep program (Vagin and Teplyakov, 2010) and a starting search model consisting of one monomer of PDB entry 1EE0. Further structural refinements utilized the Phenix suite of programs (Adams *et al.*, 2010) programs. Coot (Emsley and Cowtan, 2004) was used for graphical map inspection and manual rebuilding of atomic models.

### Example 3. *In vivo* Production using 2-Pyrone synthase Variants

*G. hybrida* 2-Pyrone Synthase (2-PS) catalyses the iterative condensation of two malonyl-CoA derived acetyl-groups with one acetyl-CoA molecule to form 6-methyl-4-hydroxy-2-pyrone (a cyclic six-carbon product), via the intermediate, acetoacetyl CoA. The atomic-resolution crystal structure of the 2-Pyrone synthase has been determined. The structure has confirmed the substantially smaller (only one third the volume) active-site cavity of 2-Pyrone synthase in comparison to the closely related enzyme, chalcone synthase (also termed 2-CHS)

(Figure 4). Notably, CHS employs a larger starter molecule, coumaryl CoA, (as opposed to the smaller acetyl CoA preferred by 2-Pyrone synthase), and performs three iterations of the acetyl-group extension (as opposed to two by 2-Pyrone synthase) to generate chalcone, a bicyclic 15-carbon product. Alfalfa CHS and 2-Pyrone synthase share 74% sequence identity and thus differ at ~100 amino-acid positions, but remarkably, substitutions in CHS at only three active-site cavity residues by the corresponding residue in 2-Pyrone synthase is sufficient to render CHS functionally identical to 2-Pyrone synthase, in terms of both substrate/product specificity and enzyme kinetics. That the functional conversion of CHS to 2-Pyrone synthase can be effected through changes at a very small subset of residues supports an intuitively simple model for the steric modulation of enzyme activity of CHS/2-Pyrone synthase, and this model sets the stage for structure-based enzyme-engineering approaches (Figure 5). The region of the 2-Pyrone synthase active site that accommodates the initial loading of the acetyl group from acetyl CoA appears to have sufficient volume to accommodate larger starter moieties, such as a benzoyl group (Figure 6). Without being bound to a particular theory, steric restriction of this “starter loading pocket” may improve the efficiency of acetyl-group loading and subsequent extension with malonyl CoA.

To this end, seven active-site residues were selected that demarcate the active-site cavity and systematically replaced with a residue possessing a longer and/or bulkier side chain (Figure 1A and 1B). Most of these substitutions proved to be deleterious for enzyme activity; however, substitutions at two positions (Leu202Phe and Leu268Met) have yielded positive results, with the mutant enzymes having an increased affinity for the reaction intermediate acetoacetyl CoA and a higher overall level of pyrone production. Active site modifications are summarized in Tables 1 and 2 below.

**Table 1. 2-Pyrone Synthase Active Site Variants with decreased activity compared to wild-type**

|                    |
|--------------------|
| C35S, I343F        |
| C35S, I343M        |
| C35S, I201M        |
| C35S, L202F, L268M |

**Table 2. 2-Pyrone Synthase Active Site Variants with increased activity compared to wild-type**

|             |
|-------------|
| C35S, T137F |
| C35S, T137L |
| C35S, I201F |
| C35S, I201L |
| C35S, L202F |
| C35S, L202M |
| C35S, L268F |
| C35S, L268M |

Positions 137, 201, 202, and 268 responded positively to mutations, whereas substitution at  
 5 position 343 consistently made the enzyme underperform.

Concurrent with enzyme-engineering approaches through active-site steric modulation, overall efficiency of pyrone production by 2-Pyrone synthase in a yeast (*Saccharomyces cerevisiae*) metabolic-engineering host was enhanced by creating variant forms of the enzyme that are more resilient and stable post expression. This component of the invention has focused  
 10 on the replacement in 2-Pyrone synthase of cysteine residues with serine or alanine (Figure 1B). Assessments of the overall production of pyrone by the mutant 2-Pyrone synthase enzymes expressed in a yeast host have demonstrated that replacements of Cys35 in particular and also Cys65, 89, 195, 346, and 372 have resulted in significant increases in pyrone yield. Without being bound to a particular theory, the C35S appears to be important for enhancing activity  
 15 and/or stability. When the C35S modification is removed, the enzyme either reverts to WT levels or underperforms. In attempts to further improve pyrone production levels, individual beneficial cysteine substitutions have been combined with multi-site cysteine substitutions. External site modifications are summarized in Tables 3 and 4 below.

**Table 3. 2-Pyrone Synthase External Site Variants with decreased activity compared to wild-type**

|                                 |
|---------------------------------|
| C35A                            |
| C35S, C65A                      |
| C35S, C346S                     |
| C35S, C135A                     |
| C65S, C89S, C372S               |
| C89S, C195S, C346A, C372S       |
| C65S, C89S, C195S, C346A, C372S |

**Table 4. 2-Pyrone Synthase External Site Variants with increased activity compared to wild-type**

|                                       |
|---------------------------------------|
| C35S                                  |
| C35S, C65S                            |
| C35S, C89S                            |
| C35S, C195S                           |
| C35S, C346A                           |
| C35S, C372A                           |
| C35S, C372S                           |
| C35S, C89S, C372S                     |
| C35S, C65S, C89S, C372S               |
| C35S, C89S, C195S, C346A, C372S       |
| C35S, C65S, C89S, C195S, C346A, C372S |

Regardless of how few cysteine substitutions are present, 2-pyrone synthase variants performed extremely well. Thus, an enzyme comprising C35S, C65S, C89S, C195S, C346A, C372S background might be useful for generating active site variants.

Measurement of the kinetic parameters ( $k_{cat}$  and  $K_m$ ) for the active-site 2-Pyrone synthase variants have shown increased levels of pyrone production (in comparison to native 2-Pyrone synthase) generated in double active-site variant (Leu202Phe and Leu268Met) when the variant 2-Pyrone synthase gene is integrated as a single copy gene in yeast. Measurements of kinetic parameters of 2-Pyrone synthase cysteine variants have also been performed. In yeast metabolic-engineering hosts, greater increases in pyrone production have been obtained with combinations of cysteine mutants than with the active-site mutants alone. The greatest increase thus far has been obtained with the Cys35Ser mutant expressed from multiply integrated copies of the 2-

Pyrone synthase gene. Such exemplary engineered enzymes may be used to prepare commercially useful chemical compounds (Figure 8).

**Example 4: 2-Pyrone synthase C35S variant showed increased triacetic lactone synthesis *in vivo*.**

Figure 14 shows the significance of 2-Pyrone synthase C35 and L268 modifications on the synthesis of triacetic lactone *in vivo*. Relative to the native *G. hybrida* 2-Pyrone synthase (WT), enzyme variants with modifications at C35 and / or L268 exhibited major changes to triacetic lactone production *in vivo*. While C35S increased titers, C35A rendered the enzyme nearly inactive, with poor synthesis. In combination with C35S, the L268F variant showed additional improvement suggesting there is active site enhancement using L268F. However, removal of C35S from the L268F variant caused triacetic lactone levels to revert to near wild-type levels. The results indicate that C35 is an important functional position outside of the catalytic cavity and that it may be important to obtain active site improvements.

**Example 5: 2-Pyrone synthase variants C35SL268F and C35ST137F showed increased triacetic lactone synthesis**

2-Pyrone synthase active site mutations coupled to the C35S mutation were screened (Figure 15). Using steric hindrance as the motivation, various residues in the active site cavity were selected for replacement with bulkier residues (M, F). Following screening *in vivo*, variants C35SL268F and C35ST137F were found to result in higher triacetic lactone levels relative to the C35S variant. Several variants showed lower levels relative to C35S, and the deleterious effect of the I201M substitution suggests side chain chemistry using methionine is unfavorable at this position. Additional reduction in triacetic lactone levels when multiple residue changes are combined indicate that changes at one position alter the impact of other neighboring residues in the cavity.

**Example 6: Variations in residues exposed to the enzyme surface increased triacetic lactone synthesis**

Individual cysteine residues, in addition to C35S, were modified to either an alanine or serine residue to mimic the size or chemistry of the thiol side chain. Significant changes in triacetic lactone titers were observed when using these cysteine variants, where residues highly

exposed to the enzyme surface provided the largest improvements (Figure 16). This screening process identified C65, C89, C135, C195, C346 and C372 as positions where substitution increases *in vivo* triacetic lactone levels.

#### 5 **Example 7: 2-Pyrone synthase C36 is important for triacetic lactone synthesis**

Following evaluation of individual surface cysteine modifications (Figure 16), different combinations of C65S, C89S, C195S, C346A, and C372S (with or without C35S) were evaluated (Figure 17). In the absence of C35S, an increase in surface cysteine substitutions resulted in reduced triacetic lactone levels. In contrast, variants incorporating C35S showed  
10 comparable levels to the single C35S variant. However, no additional increase in *in vivo* TAL levels was observed beyond the C35SC372S double variant. These results further validate the importance of the C35 position, possibly in enzyme stability and protection from *in vivo* degradative processes.

#### 15 **Example 8: C35SC372S protects 2-Pyrone synthase from proteolysis**

The C35S and C35SC372S 2-PS variants were evaluated in strains with protease knockouts. Knocking out *PEP4* or *PRB1* significantly improved triacetic lactone levels when the wild-type 2-Pyrone synthase was used (Figures 7 and 18). However, the increase was smaller with the C35S variant, and no improvement was seen with C35SC372S. These results suggest  
20 that C35SC372S is sufficient to protect the synthase from yeast proteolytic mechanisms targeting the enzyme.

The results described herein were obtained using the following materials and methods.

#### **Strains and plasmids**

25 *Escherichia coli* strain XL1-Blue (Stratagene, Santa Clara, CA) was used for amplification of plasmids. *S. cerevisiae* strain BJ5464 (Jones, 1991) was used as the base strain for TAL production. The engineered strains BJ $\Delta$ *pyc2* $\Delta$ *nte1* and protease limited strains BYt $\Delta$ *pep4*, BYt $\Delta$ *prb1*, and BYt $\Delta$ *pep4* $\Delta$ *prb1* (Cardenas and Da Silva, 2014) were used for the *in vivo* 2-PS evaluation studies. *S. cerevisiae* strain BY4741 $\Delta$ *trp1* (Open Biosystems, Huntsville,  
30 AL) was also used for comparison as a control.

The high copy 2 $\mu$ -based pXP842 vector (Shen et al., 2012), harboring the glucose-repressed *ADH2* promoter and the *CYC1* terminator with a loxP-flanked *URA3* selection marker, was used to express all 2-PS enzyme variants. Yeast cells were transformed as previously described (Gietz et al., 1992; Hill et al., 1991) using selective SDC-A plates. Cell colonies were  
5 allowed to grow for 3-5 days before generating inoculum cultures or glycerol stocks for long-term storage.

The *Gerbera hybrida* *g2ps1* gene (encoding 2-PS) and all variant genes were PCR amplified from the pHIS8 cassette (Jez et al., 2000). Following *SpeI* and *XhoI* digestion, the gene was inserted into pXP842 (Shen et al., 2012) using the Rapid DNA Ligation Kit (Thermo  
10 Scientific, Waltham, MA). Plasmid recovery was performed using the GeneJet<sup>TM</sup> Plasmid Miniprep Kit (Thermo Scientific, Waltham, MA) and DNA sequence analysis confirmed the correct sequence of all PCR-amplified inserts (GeneWiz, South Plainfield, NJ; Eton Biosciences, San Diego, CA).

The KOD Hot-start polymerase (EMD Chemicals, San Diego, CA) was used in PCR  
15 reactions for plasmid constructs. Restriction enzymes, T4 DNA ligase, Taq DNA polymerase, and deoxynucleotides were purchased from New England Biolabs. Oligonucleotide primers were purchased from IDT DNA (San Diego, CA).

### Media and cultivation

Luria-Bertani (LB) media was used for proliferation of XL1-Blue cells with 150 mg/L  
20 ampicillin for selection of plasmid-containing strains (Sambrook and Russell, 2001). For inoculum cultures, *S. cerevisiae* strains were grown for 16 h in 5 mL selective SDC(A) medium (1% dextrose, 0.67% yeast nitrogen base, 0.5% Bacto casamino acids, 0.5% ammonium sulfate and 100 mg/L adenine) in an air shaker (New Brunswick Scientific) at 250 rpm and 30°C. These were used to inoculate cultures in complex YPD media (1% dextrose; 1% Bacto yeast extract;  
25 2% Bacto peptone) to an initial OD<sub>600</sub> of 0.3 (Shimadzu UV-2450 UV-VIS Spectrophotometer, Columbia, MD). At 48 h, samples were taken and cell density determined. The samples were then centrifuged at 3,000 rpm (2,600 g) for 5 min at 4°C (Beckman GS-6R Centrifuge, Brea, CA), and the supernatants stored at 4°C for HPLC analysis of TAL levels in the culture broth.

### HPLC Assay

The concentration of triacetic acid lactone was measured by HPLC using a Shimadzu HPLC system: LC-10AT pumps (Shimadzu), UV-VIS detector (SPD-10A VP, Shimadzu), Zorbax SB-C18 reversed-phase column (2.1x150 mm, Agilent Technologies). Acetonitrile buffered in 1% acetic acid was used as the mobile phase, while HPLC grade water buffered in 1% acetic acid was used as the aqueous phase. A gradient program using a 95-85% Pump B gradient (H<sub>2</sub>O with 1% acetic acid) provided an elution time of approximately 12 minutes (flow rate 0.25 mL/min, column temperature 25°C).

### Pyrrone Synthase Assay

Activities of 2-Pyrrone synthase and variants (0.5-2.5 g) were determined by monitoring product formation using a LC/MS assay. Standard assay conditions were 100 mM HEPES (pH 7.0), 30 M starter-CoA (usually acetyl-CoA), and 60 M malonyl-CoA in 2.0 ml at 25°C. Reactions were quenched with 5% acetic acid and extracted with ethyl acetate. Product formation was determined by analysis of extracts on a Hewlett-Packard HP1100 MSD single quadrupole mass spectrometer coupled to a Zorbax SB-C18 column (5 m, 2.1 mm x 150 mm). High performance liquid chromatography conditions were as follows: gradient system from 0 to 100% methanol in water (each containing 0.2% acetic acid) over 20 min; flow rate 0.25 ml/min. LC/MS/MS data: 2PY (6-methyl-4-hydroxy-2-pyrrone), Rt = 9.62 min. Steady-state kinetic constants were determined from initial velocity measurements, in which product formation was linear over the time periods monitored, using standard assay conditions with a fixed malonyl-CoA concentration (120 M) and varying acetyl-CoA concentrations (0.5-50 M).

### Thermal-shift Binding Assay

The shifts in melting temperature of 2-Pyrrone synthase proteins induced by a series of coenzyme-A ligands at varying concentrations were measured using a ThermoFluor-type assay similar to one previously published (Niesen *et al.*, 2007). Protein melting was monitored using a LightCycler480 System II (Roche), with the following temperature-ramping program: 30 seconds at 20°C, ramp up to 85°C at 0.06°C/second, 30 seconds at 20°C. Using SYPRO Orange (Sigma; excitation wavelength 483 nm, emission wavelength 568 nm), an environmentally-sensitive fluorescent dye that can interact with hydrophobic amino-acid residues, the thermal

melting of a protein can be monitored through the increased fluorescence signal accompanying exposure to the solvent medium of the hydrophobic core residues of a denatured protein. The maximum of the first-derivative curve of the fluorescence profile is typically used as a measure of the melting temperature ( $T_m$ ).

5

## References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.D.,  
10 Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D Biol. Crystallogr.* 66, 213-221.
- Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM:  
15 a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* 67, 271-281.
- Cardenas, J., Da Silva, N.A., 2014. Metabolic engineering of *Saccharomyces cerevisiae* for the production of triacetic acid lactone. *Metabolic Engineering* 25, 194–203.
- 20 Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Evans, P.R. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62, 72-82.
- 25 Gietz, D., Jean, A.S., Woods, R.A., Schiestl, R.H., 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucl. Acids Res.* 20, 1425–1425
- Hill, J., Donald, K.A.I.G., Griffiths, D.E., 1991. DMSO-enhanced whole cell yeast transformation. *Nucl. Acids Res.* 19, 5791–5791
- 30 Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P. (2000a). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* 7, 786-791.
- 35 Jez, J.M., Austin, M.B., Ferrer, J., Bowman, M.E., Schroder, J., and Noel, J.P. (2000b). Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.* 7, 919-930.
- 40 Niesen, F. H., Berglund, H. and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protocols* 9, 2212-2221.
- Sambrook, J.J., Russell, D.D.W., 2001. *Molecular cloning: a laboratory manual*. Vol. 2. CSHL Press.

Shen, M.W.Y., Fang, F., Sandmeyer, S., Da Silva, N.A., 2012. Development and characterization of a vector set with regulated promoters for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast* 29, 495–503.

5

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr.* 66, 22-25.

### Other Embodiments

10 From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed  
15 elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference.

## CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule encoding a polyketide synthase variant comprising an altered amino acid residue corresponding to amino acid positions C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.

2. The isolated nucleic acid molecule of claim 1, wherein the polyketide synthase variant is 2-pyrone synthase.

3. The isolated nucleic acid molecule of claim 1, wherein the altered amino acid residue is selected from the group consisting of C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M.

4. The isolated nucleic acid molecule of claim 2, wherein the altered amino acid residue is C35S.

5. The isolated nucleic acid molecule of claim 2, wherein the encoded polyketide synthase variant comprises a combination of altered amino acid residues selected from the group consisting of C35S, T137F; C35S, T137L; C35S, I201F; C35S, I201L; C35S, L202F; C35S, L202M; C35S, L268F; C35S, L268M; C35S; C35S, C65S; C35S, C89S; C35S, C195S; C35S, C346A; C35S, C372A; C35S, C372S; C35S, C89S, C372S; C35S, C65S, C89S, C372S; C35S, C89S, C195S, C346A, C372S; and C35S, C65S, C89S, C195S, C346A, and C372S.

6. The isolated nucleic acid molecule of claim 2, wherein the encoded polyketide synthase variant comprises C35SC372S.

7. The isolated nucleic acid molecule of claim 1, wherein the activity and/or stability of the polyketide synthase is increased relative to a reference polypeptide.

8. The isolated nucleic acid molecule of claim 7, wherein the activity comprises production of a compound selected from the group consisting of a pyrone, a chromone, a lactone, a polyhydroxynaphthalene, a phloroglucinol, a resorcinol, a resorcinol acid, SEK 4, SEK4b, aloesone, and combinations thereof.
9. The isolated nucleic acid molecule of claim 8, wherein the lactone is triacetic acid lactone.
10. A vector comprising the isolated nucleic acid molecule of any one of claims 1-9.
11. A cell comprising the isolated nucleic acid molecule of any one of claims 1-9 and/or the vector of claim 10.
12. The cell of claim 11, wherein the cell is a yeast, plant, algae, bacterial, mammalian, or insect cell.
13. The cell of claim 12, wherein the cell is *in vitro* or *in vivo*.
14. A plant comprising the isolated nucleic acid molecule of any one of claims 1-9 and/or the vector of claim 10.
15. The plant of claim 14, wherein the nucleic acid molecule is expressed in a glandular tricone.
16. An isolated polyketide synthase polypeptide comprising an altered amino acid residue corresponding to amino acid positions C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.
17. The isolated polyketide synthase polypeptide of claim 16, wherein the polyketide synthase polypeptide is 2-pyrone synthase.

18. The isolated polyketide synthase polypeptide of claim 17, wherein the altered amino acid residue is selected from the group consisting of C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M.

5

19. The isolated polyketide synthase polypeptide of claim 16, wherein the altered amino acid residue is C35S.

20. The isolated polyketide synthase polypeptide of claim 16, wherein the encoded polyketide synthase variant comprises a combination of altered amino acid residues selected from the group consisting of C35S, T137F; C35S, T137L; C35S, I201F; C35S, I201L; C35S, L202F; C35S, L202M; C35S, L268F; C35S, L268M; C35S; C35S, C65S; C35S, C89S; C35S, C195S; C35S, C346A; C35S, C372A; C35S, C372S; C35S, C89S, C372S; C35S, C65S, C89S, C372S; C35S, C89S, C195S, C346A, C372S; and C35S, C65S, C89S, C195S, C346A, and C372S.

15

21. The isolated polyketide synthase of claim 16, wherein the polyketide synthase polypeptide comprises C35SC372S.

22. The isolated polyketide synthase of claim 16, wherein the polyketide synthase polypeptide comprises C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M.

20

23. The isolated polyketide synthase of any of claims 16-21, wherein the activity and/or stability of the synthase is increased relative to a wild-type 2-pyrone synthase polypeptide.

25

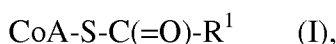
24. The isolated polyketide synthase of claim 23, wherein the activity comprises production of a compound selected from the group consisting of a pyrone, a chromone, a lactone, a polyhydroxynaphthalene, a phloroglucinol, a resorcinol, a resorcinol acid, SEK 4, SEK4b, alonesone, and any combinations thereof.

30

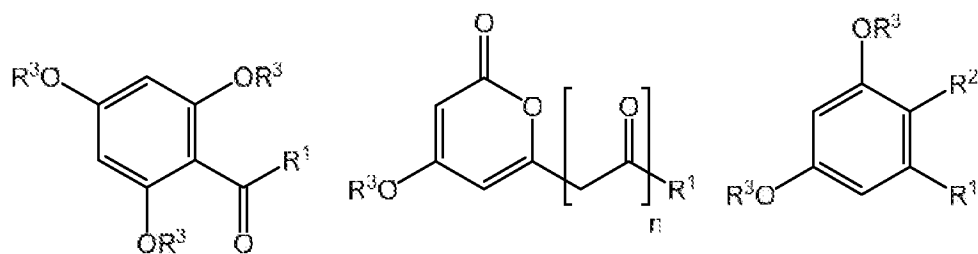
25. The isolated polyketide synthase of claim 24, wherein the lactone comprises triacetic acid lactone.
26. A method of increasing the activity of a polyketide synthase, the method comprising  
5 introducing an altered amino acid residue at a position corresponding to C35, T137, I201, L202, L268, C65, C89, C195, C346; and/or C372 of 2-pyrone synthase.
27. The method of claim 26, wherein the polyketide synthase is 2-pyrone synthase.
- 10 28. The method of claim 26, wherein the altered amino acid residue is selected from the group consisting of C35S, C65S, C89S, C195S, C346A, C372S, L202F, and L268M.
29. The method of claim 26, wherein the altered amino acid residue is C35S.
- 15 30. The method of claim 26, wherein  $k_{cat}$  of the synthase is increased and/or  $K_m$  of the synthase is decreased relative to a wild-type 2-pyrone synthase polypeptide.
31. The method of claim 26, wherein the activity comprises production of a compound selected from the group consisting of a pyrone, a chromone, a lactone, a  
20 polyhydroxynaphthalene, a phloroglucinol, a resorcinol, a resorcinol acid, SEK 4, SEK4b, alonesone, and any combinations thereof.
32. The method of claim 31, wherein the lactone comprises triacetic acid lactone.
- 25 33. The method of claim 26, wherein the variant polyketide synthase is expressed in a yeast, plant, algae, bacterial, mammalian, or insect cell.
34. The method of claim 27, wherein the cell is *in vitro* or *in vivo*.
- 30 35. A method of preparing a polyketide using the cell of claims 11-13.

36. The method of claim 33 wherein the polyketide is a pyrone, a chromone, a lactone, a polyhydroxynaphthalene, a phloroglucinol, a resorcinol, a resorcinol acid, SEK 4, SEK4b, alonesone, and any combinations thereof.

37. A method of preparing a cyclic compound, or a salt, solvate or derivative thereof, the method comprising contacting the isolated polyketide synthase of any one of claims 15-23 with a substrate of formula (I), or a salt, solvate or derivative thereof:



wherein the cyclic compound is selected from the group consisting of:



, wherein:

R<sup>1</sup> is selected from the group consisting of C<sub>1</sub>-C<sub>20</sub> alkyl, C<sub>1</sub>-C<sub>20</sub> cycloalkyl, aryl-(C<sub>1</sub>-C<sub>20</sub> alkyl), carboxy-(C<sub>1</sub>-C<sub>20</sub> alkyl), aryl, heterocyclyl, heteroaryl, aryl-CH=CH-, heterocyclyl-CH=CH- and heteroaryl-CH=CH-,

R<sup>2</sup> is selected from the group consisting of H, carboxy, C(=O)O(C<sub>1</sub>-C<sub>6</sub>)alkyl, cyano, -C(=O)NH<sub>2</sub>, -C(=O)NH(C<sub>1</sub>-C<sub>6</sub>)alkyl, and -C(=O)N((C<sub>1</sub>-C<sub>6</sub>)alkyl)<sub>2</sub>;

each occurrence of R<sup>3</sup> is independently selected from the group consisting of H, C<sub>1</sub>-C<sub>20</sub> alkyl, aryl-(C<sub>1</sub>-C<sub>20</sub> alkyl), C<sub>1</sub>-C<sub>20</sub> acyl, aroyl, and heteroaroyl; and, n is 0, 1, 2 or 3;

wherein each alkyl, cycloalkyl, aryl, heterocyclyl or heteroaryl group is independently optionally substituted with at least one substituent selected from the group consisting of F, Cl, Br, I, hydroxy, alkoxy, amino, nitro, monoalkylamino, dialkylamino, carboxy, C(=O)O(C<sub>1</sub>-C<sub>6</sub>)alkyl, trifluoromethyl, cyano, -C(=O)NH<sub>2</sub>, -C(=O)NH(C<sub>1</sub>-C<sub>6</sub>)alkyl, -C(=O)N((C<sub>1</sub>-C<sub>6</sub>)alkyl)<sub>2</sub>, -SO<sub>2</sub>NH<sub>2</sub>, and -C(=NH)NH<sub>2</sub>.

38. The method of claim 37, wherein R<sup>1</sup> is selected from the group consisting of C<sub>1</sub>-C<sub>20</sub> alkyl, carboxy-(C<sub>1</sub>-C<sub>20</sub> alkyl), aryl, heteroaryl, aryl-CH=CH-, and heteroaryl-CH=CH-.

39. The method of claim 37, wherein  $R^1$  is selected from the group consisting of  $C_1$ - $C_{20}$  alkyl, carboxy-( $C_1$ - $C_{20}$  alkyl), aryl, and aryl-CH=CH-.

5 40. The method of claim 37, wherein each occurrence of aryl comprises phenyl.

41. A kit comprising an isolated polyketide synthase, a nucleic acid molecule, a vector or a cell of any preceding claim.

1/150

FIG. 1A

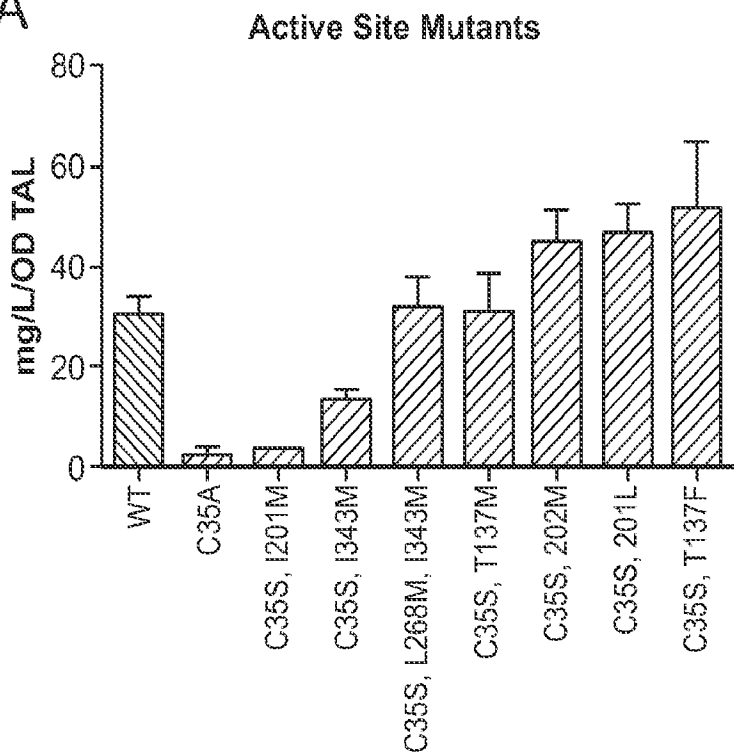
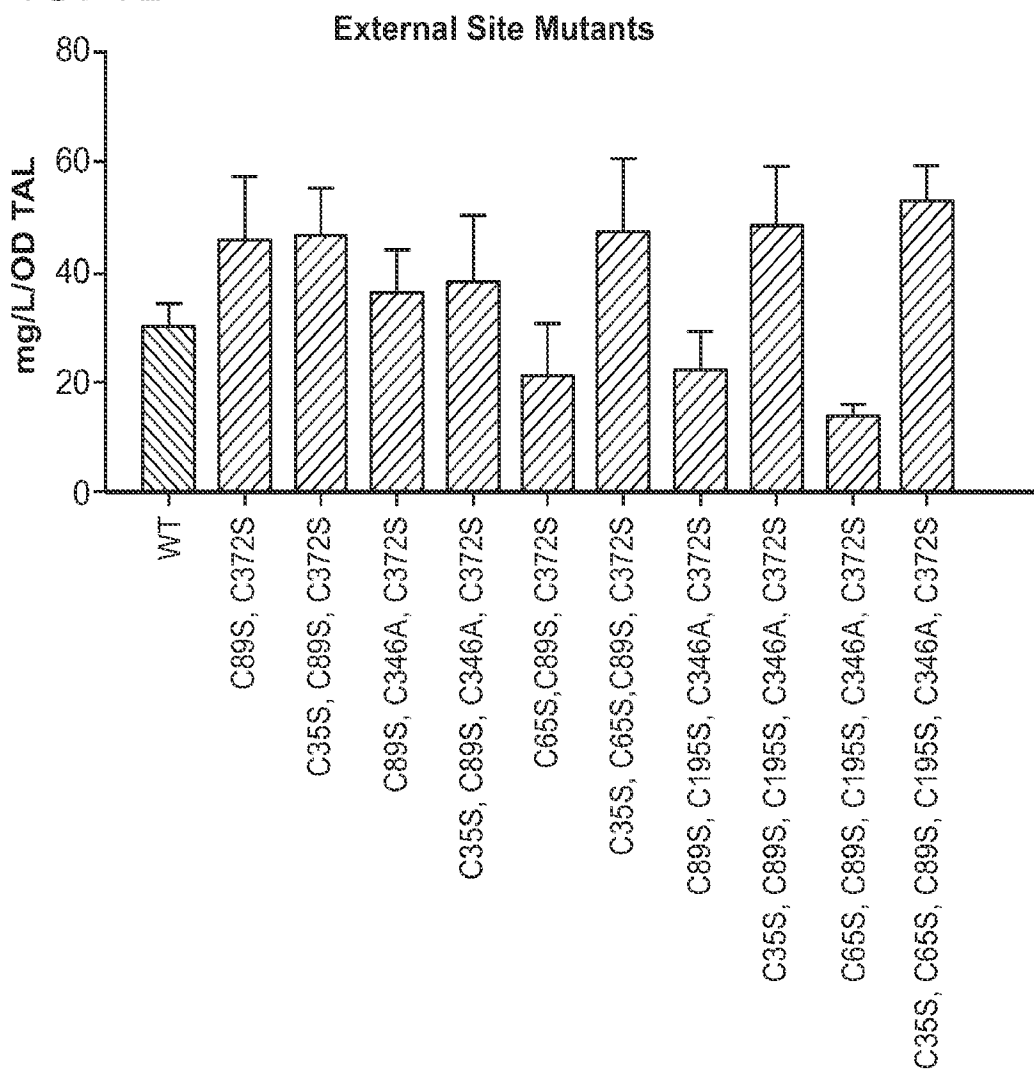


FIG. 1B



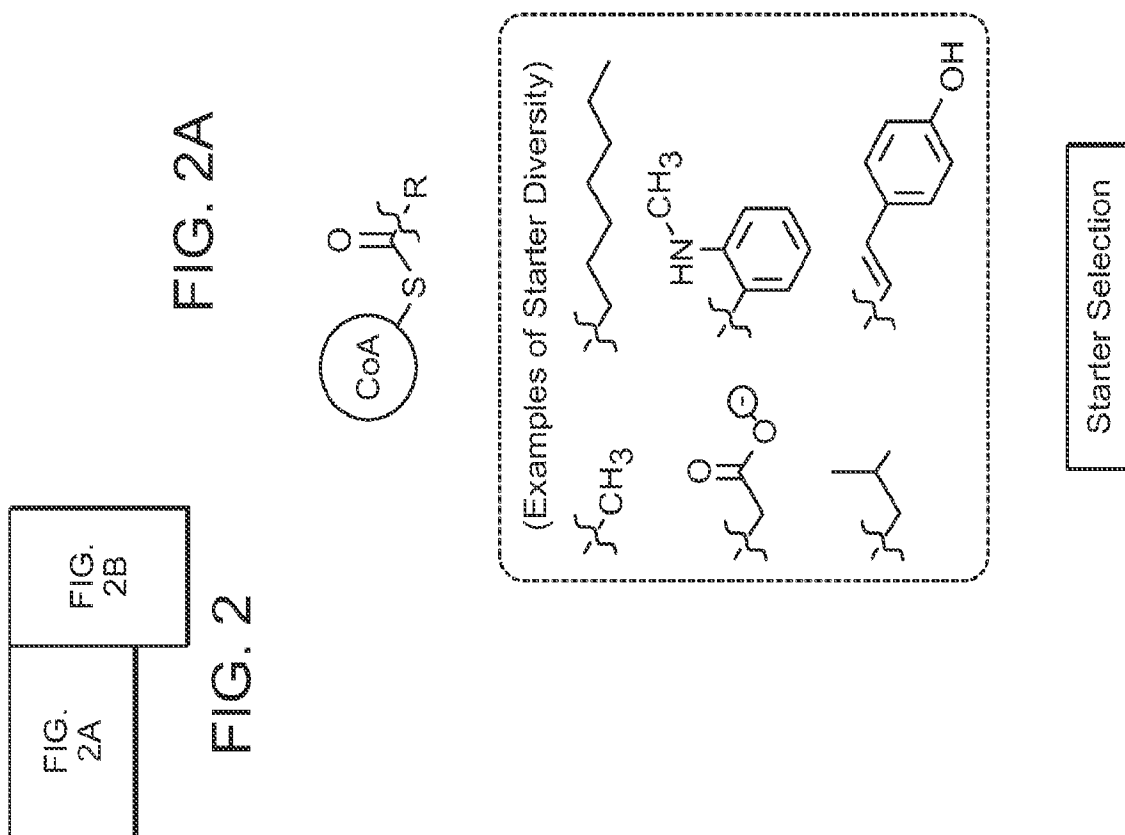
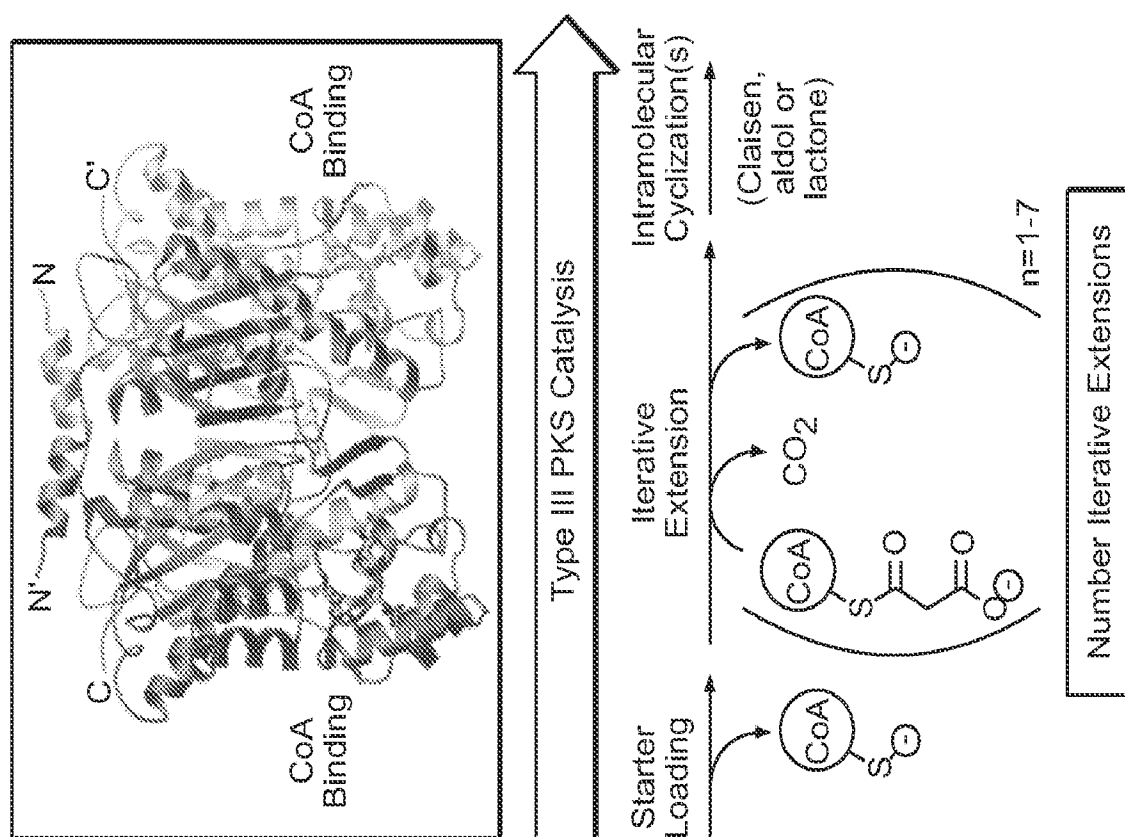
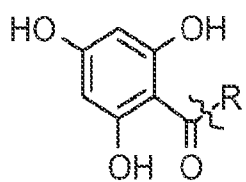
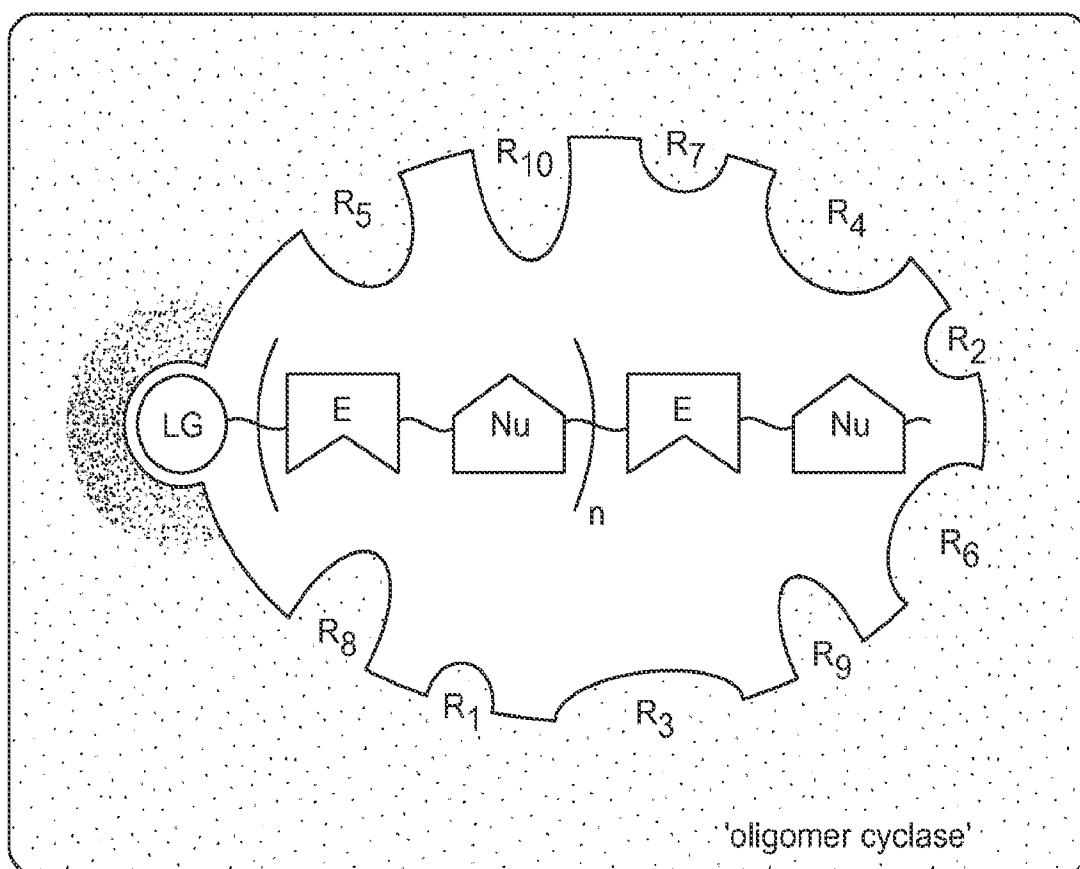
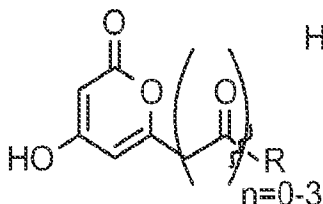


FIG. 2B

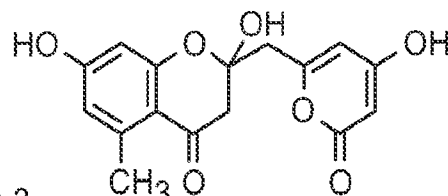
3/150



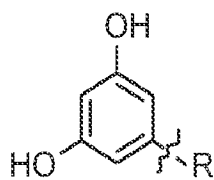
phloroglucinols



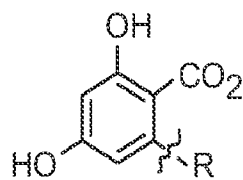
lactones



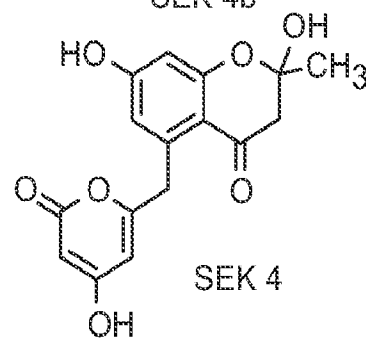
SEK 4b



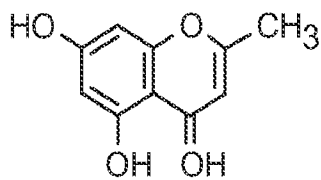
resorcinols



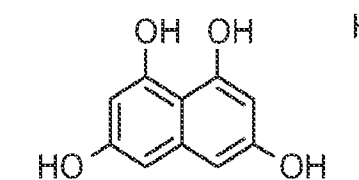
resorcinolic acids?



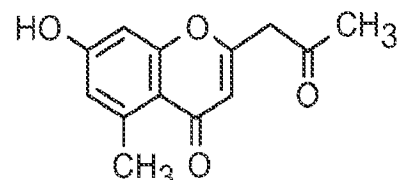
SEK 4



chromone

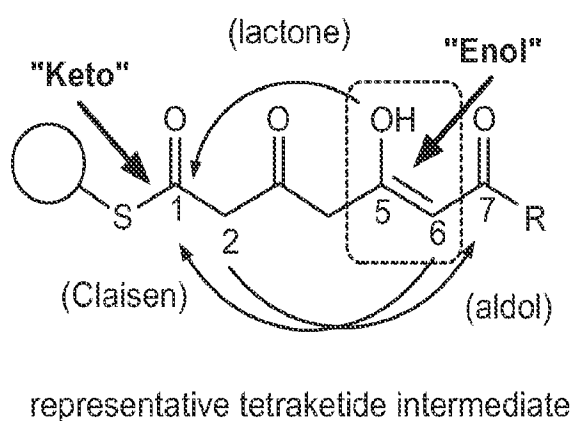


tetrahydroxynaphthalene



aloesone

Cyclization Mechanism



Tautomeric  
States and  
Atomic  
Proximity  
Modulate  
Bond  
Formation

repetitive & flexible linear oligomers...

activating Leaving Groups (chemical handles)...

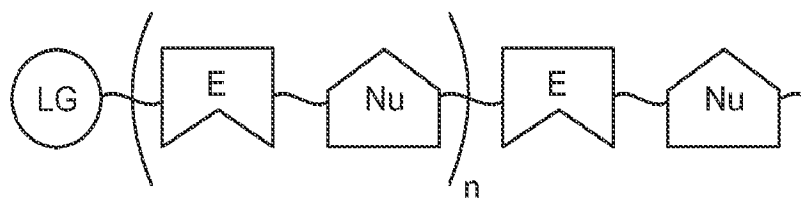


FIG. 3

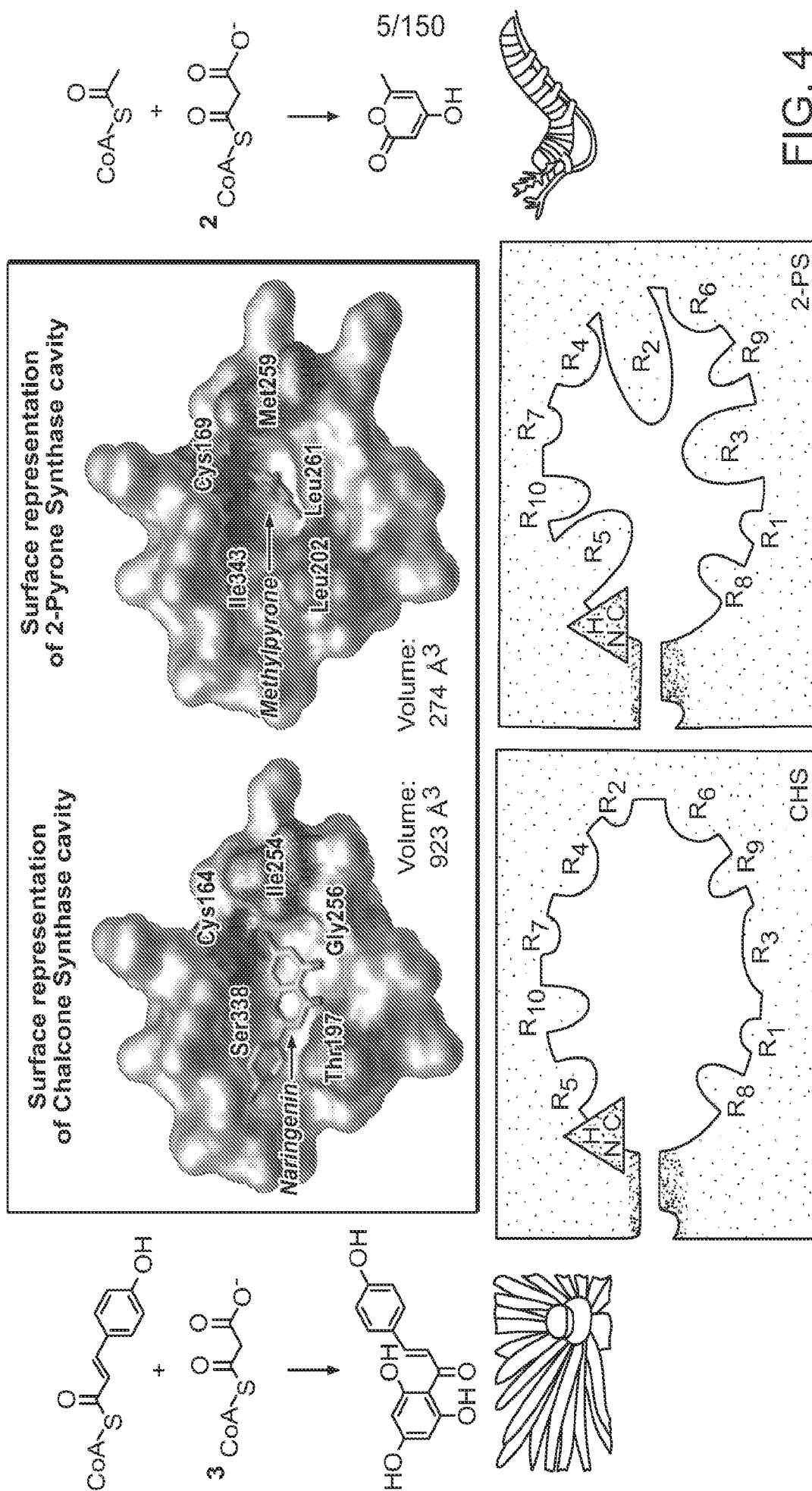
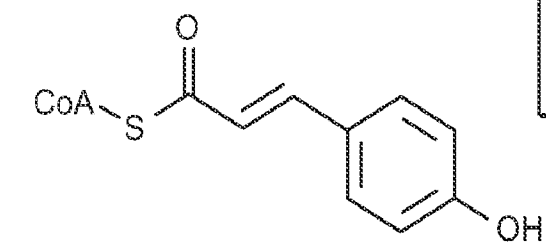
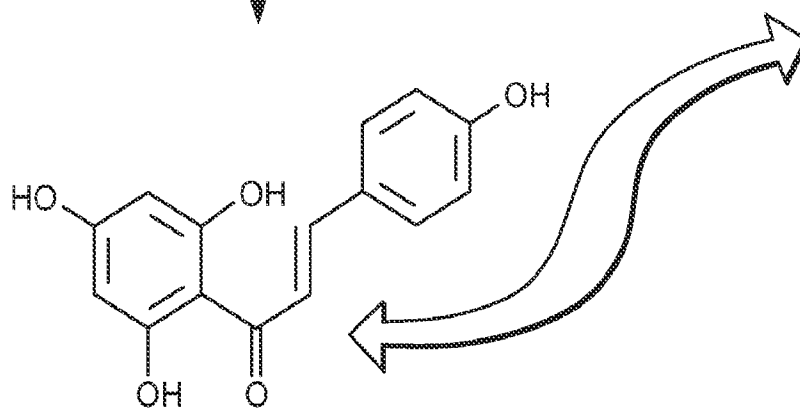
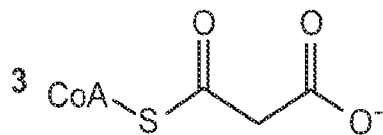


FIG. 4

6/150



+



|            |            |            |
|------------|------------|------------|
| FIG.<br>5A | FIG.<br>5B | FIG.<br>5C |
|------------|------------|------------|

FIG. 5

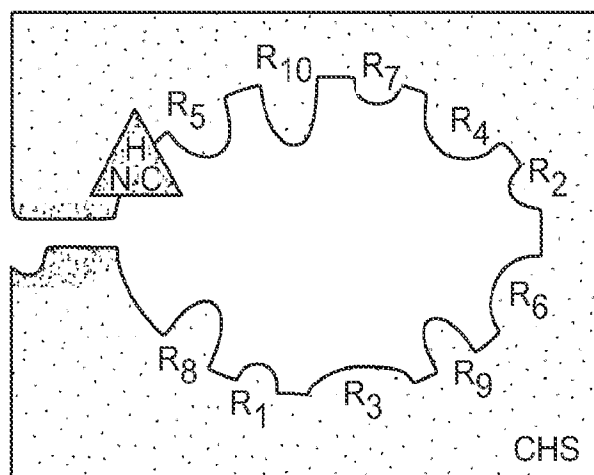


FIG. 5A

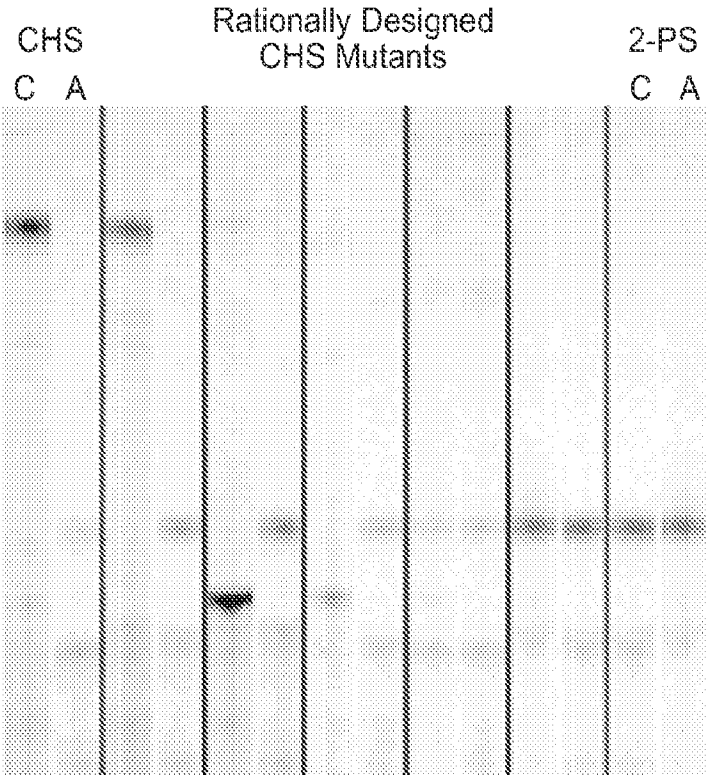


FIG. 5B

8/150

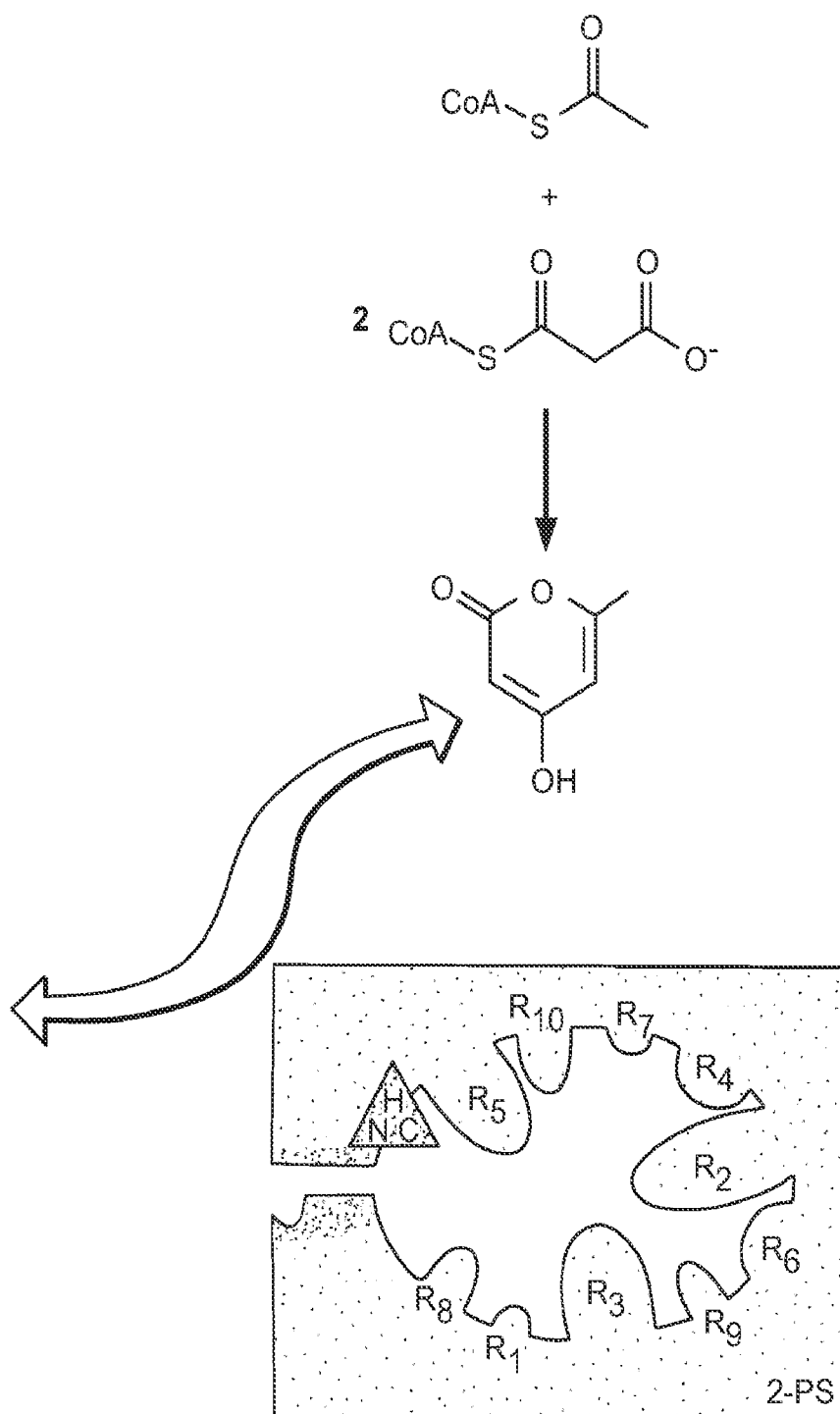
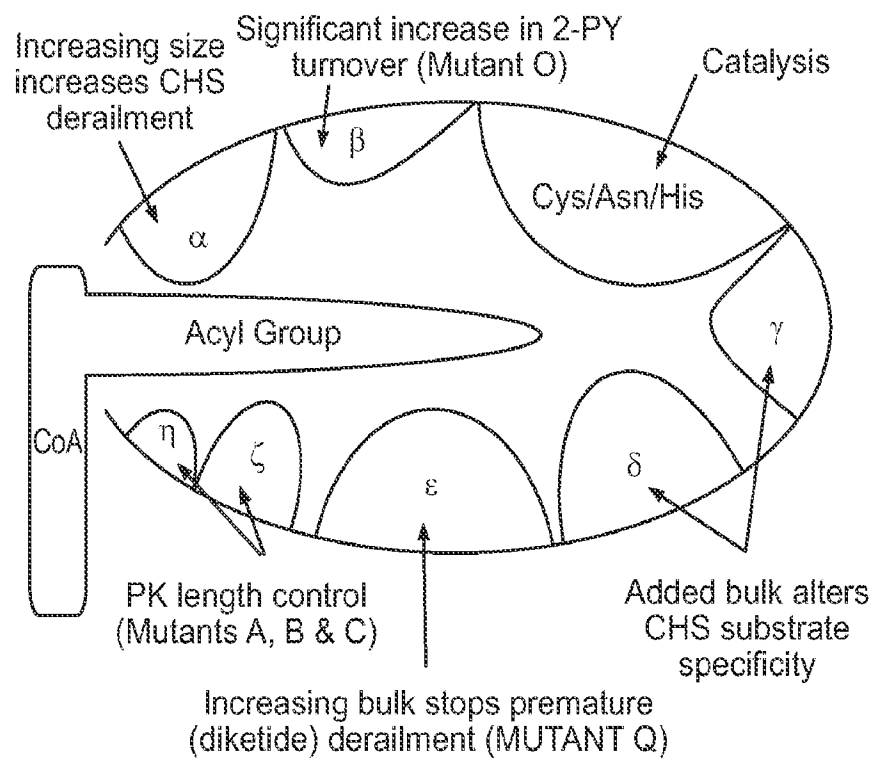


FIG. 5C

**2-Pyrone Synthase Active Site****FIG. 6**

10/150

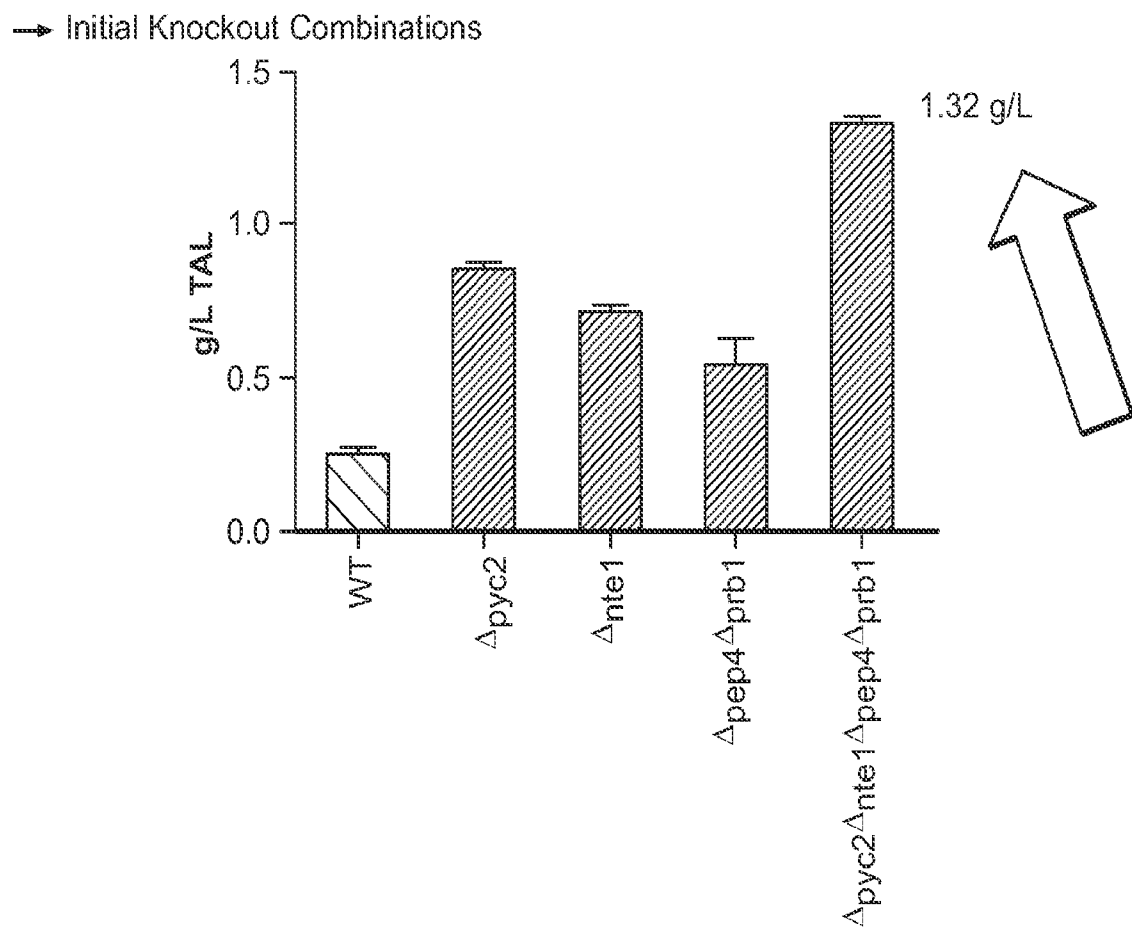
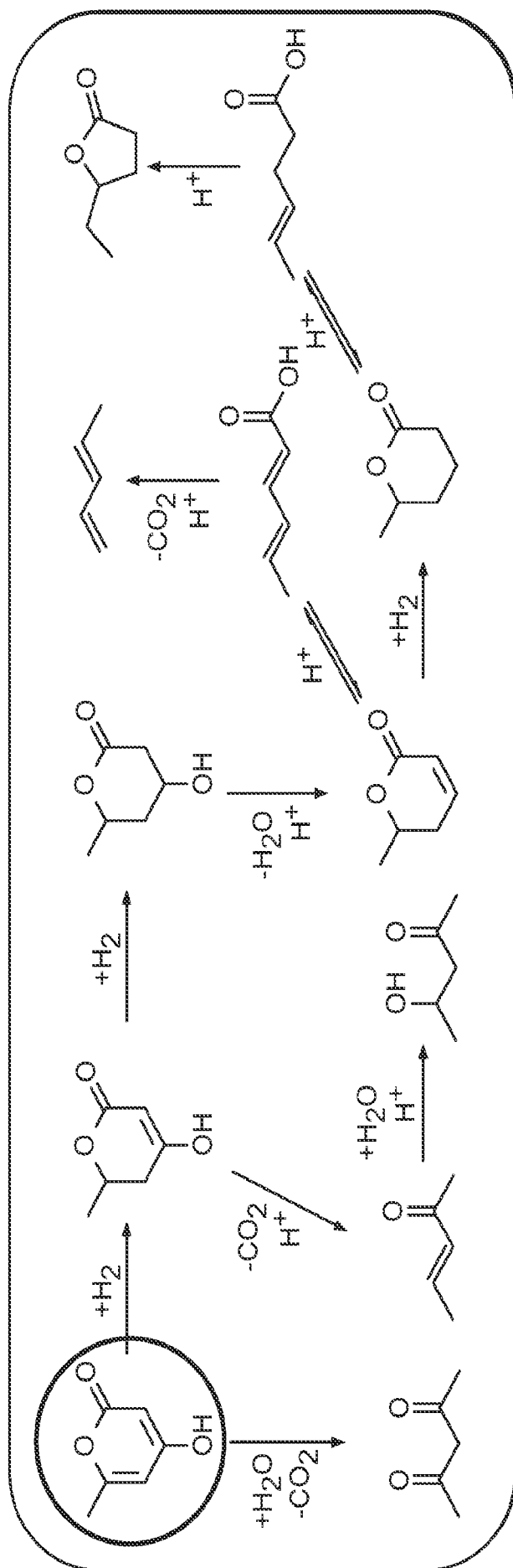


FIG. 7



11 potential chemical intermediates and end-products identified and synthesized from TAL

Pyrones are promising biorenewable platform chemicals

M. Chia, T.J. Schwartz, B.H. Shanks, J.A. Dumesic, *Green Chemistry*, **14**, 1850-1853, 2012

FIG. 8

## FIG. 9

|                 |                   |     |             |         |         |
|-----------------|-------------------|-----|-------------|---------|---------|
| > 2-PS variant= | 1 Substitutions=  | 0:  | (wild type) |         |         |
| > 2-PS variant= | 2 Substitutions=  | 1:  | (C169A)     |         |         |
| > 2-PS variant= | 3 Substitutions=  | 1:  | (C 35A)     |         |         |
| > 2-PS variant= | 4 Substitutions=  | 1:  | (L202F)     |         |         |
| > 2-PS variant= | 5 Substitutions=  | 1:  | (L268F)     |         |         |
| > 2-PS variant= | 6 Substitutions=  | 1:  | (L268M)     |         |         |
| > 2-PS variant= | 7 Substitutions=  | 2:  | (L202F)     | (L268M) |         |
| > 2-PS variant= | 8 Substitutions=  | 2:  | (C 89S)     | (C372S) |         |
| > 2-PS variant= | 9 Substitutions=  | 3:  | (C 89S)     | (C346A) | (C372S) |
| > 2-PS variant= | 10 Substitutions= | 3:  | (C 65S)     | (C 89S) | (C372S) |
| > 2-PS variant= | 11 Substitutions= | 4:  | (C 89S)     | (C195S) | (C372S) |
| > 2-PS variant= | 12 Substitutions= | 5:  | (C 65S)     | (C 89S) | (C346A) |
| > 2-PS variant= | 13 Substitutions= | 8:  | (C 35S)     | (C 65S) | (C195S) |
| (C346A) (C372S) |                   |     |             | (C 89S) | (C372S) |
| > 2-PS variant= | 14 Substitutions= | 10: | (C 35S)     | (C 89S) | (C195S) |
| (M259L) (L261N) | (C346A) (C372S)   |     |             | (T137F) | (L202G) |
| > 2-PS variant= | 15 Substitutions= | 10: | (C 35S)     | (C 89S) | (L202G) |
| (M259L) (L261N) | (C346A) (C372S)   |     |             | (C195S) | (I201F) |
| > 2-PS variant= | 16 Substitutions= | 10: | (C 35S)     | (C 89S) | (L202G) |
| (M259L) (L261N) | (C346A) (C372S)   |     |             | (T137F) | (C195S) |
| > 2-PS variant= | 17 Substitutions= | 10: | (C 35S)     | (C 89S) | (L202G) |
| (L261N) (L268M) | (C346A) (C372S)   |     |             | (C195S) | (M259L) |
| > 2-PS variant= | 18 Substitutions= | 3:  | (C 35S)     | (C 89S) |         |
| > 2-PS variant= | 19 Substitutions= | 4:  | (C 35S)     | (C 89S) | (C372S) |

12/150

## FIG. 9 (continued)

|                         |    |                |    |         |         |         |                 |
|-------------------------|----|----------------|----|---------|---------|---------|-----------------|
| > 2-PS variant=         | 20 | Substitutions= | 4: | (C 35S) | (C 65S) | (C 89S) | (C372S)         |
| > 2-PS variant=         | 21 | Substitutions= | 5: | (C 35S) | (C 89S) | (C195S) | (C346A)         |
| > 2-PS variant=         | 22 | Substitutions= | 6: | (C 35S) | (C 65S) | (C 89S) | (C195S)         |
| > 2-PS variant=         | 23 | Substitutions= | 2: | (C 35S) | (T137F) |         | (C372S)         |
| > 2-PS variant=         | 24 | Substitutions= | 2: | (C 35S) | (T137M) |         |                 |
| > 2-PS variant=         | 25 | Substitutions= | 2: | (C 35S) | (M259F) |         |                 |
| > 2-PS variant=         | 26 | Substitutions= | 2: | (C 35S) | (T137L) |         |                 |
| > 2-PS variant=         | 27 | Substitutions= | 2: | (C 35S) | (I343F) |         |                 |
| > 2-PS variant=         | 28 | Substitutions= | 2: | (C 35S) | (I343M) |         |                 |
| > 2-PS variant=         | 29 | Substitutions= | 2: | (C 35S) | (I201F) |         |                 |
| > 2-PS variant=         | 30 | Substitutions= | 2: | (C 35S) | (C169A) |         |                 |
| > 2-PS variant=         | 31 | Substitutions= | 2: | (C 35S) | (L268M) |         |                 |
| > 2-PS variant=         | 32 | Substitutions= | 2: | (C 35S) | (L202F) |         |                 |
| > 2-PS variant=         | 33 | Substitutions= | 2: | (C 35S) | (I201M) |         |                 |
| > 2-PS variant=         | 34 | Substitutions= | 2: | (C 35S) | (I201L) |         |                 |
| > 2-PS variant=         | 35 | Substitutions= | 2: | (C 35S) | (L202M) |         |                 |
| > 2-PS variant=         | 36 | Substitutions= | 4: | (C 35S) | (L202G) | (M259L) | (L261N)         |
| > 2-PS variant=         | 37 | Substitutions= | 9: | (C 35S) | (C 65S) | (C 89S) | (C195S)         |
| (L261N) (C346A) (C372S) |    |                |    |         |         |         | (L202G) (M259L) |
| > 2-PS variant=         | 38 | Substitutions= | 2: | (C 35S) | (L202G) |         |                 |
| > 2-PS variant=         | 39 | Substitutions= | 3: | (C 35S) | (M259L) | (L261N) |                 |
| > 2-PS variant=         | 40 | Substitutions= | 7: | (C 35S) | (C 65S) | (C 89S) | (C195S)         |
| (C372S)                 |    |                |    |         |         |         | (L202G) (C346A) |

## FIG. 9 (continued)

|                                   |    |         |         |         |         |         |
|-----------------------------------|----|---------|---------|---------|---------|---------|
| > 2-PS variant= 41 Substitutions= | 1: | (C 35S) |         |         |         |         |
| > 2-PS variant= 42 Substitutions= | 4: | (C 35S) | (T137L) | (L202F) | (L268M) |         |
| > 2-PS variant= 43 Substitutions= | 5: | (C 35S) | (T137M) | (L202F) | (L268M) | (I343F) |
| > 2-PS variant= 44 Substitutions= | 4: | (C 35S) | (L202F) | (L268M) | (I343F) |         |
| > 2-PS variant= 45 Substitutions= | 2: | (C 35S) | (C372S) |         |         |         |
| > 2-PS variant= 46 Substitutions= | 2: | (C 35S) | (C372A) |         |         |         |
| > 2-PS variant= 47 Substitutions= | 2: | (C 35S) | (C346S) |         |         |         |
| > 2-PS variant= 48 Substitutions= | 2: | (C 35S) | (C346A) |         |         |         |
| > 2-PS variant= 49 Substitutions= | 2: | (C 35S) | (C195S) |         |         |         |
| > 2-PS variant= 50 Substitutions= | 2: | (C 35S) | (C195A) |         |         |         |
| > 2-PS variant= 51 Substitutions= | 2: | (C 35S) | (C135S) |         |         |         |
| > 2-PS variant= 52 Substitutions= | 2: | (C 35S) | (C135A) |         |         |         |
| > 2-PS variant= 53 Substitutions= | 2: | (C 35S) | (C 89S) |         |         |         |
| > 2-PS variant= 54 Substitutions= | 2: | (C 35S) | (C 89A) |         |         |         |
| > 2-PS variant= 55 Substitutions= | 2: | (C 35S) | (C 65S) |         |         |         |
| > 2-PS variant= 56 Substitutions= | 2: | (C 35S) | (C 65A) |         |         |         |
| > 2-PS variant= 57 Substitutions= | 2: | (C 35S) | (C169S) |         |         |         |
| > 2-PS variant= 58 Substitutions= | 2: | (C 35S) | (N341H) |         |         |         |
| > 2-PS variant= 59 Substitutions= | 2: | (C 35S) | (H308Q) |         |         |         |
| > 2-PS variant= 60 Substitutions= | 3: | (C 35S) | (L202F) | (L268F) |         |         |
| > 2-PS variant= 61 Substitutions= | 3: | (C 35S) | (L268M) | (I343M) |         |         |
| > 2-PS variant= 62 Substitutions= | 3: | (C 35S) | (L202F) | (L268M) |         |         |

15/150

## FIG. 10

```

> 2-PS variant= 2 Substitutions= 1: (C169A)
MGYSDDVEVIREAGRAQGLATIIAIGTATPPNCVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQQaAAGGTVLRLAKDLAENNKGSRVLI VCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTI LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSP LGITDWN SVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLSEYGNLISACVLFIIDEVKRKRSM AEGKSTTGEGLDCGVLF GFGPGMTVETVVLRSVRVTA AAVAN
GN

> 2-PS variant= 3 Substitutions= 1: (C 35A)
MGYSDDVEVIREAGRAQGLATIIAIGTATPPNaVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI VCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTI LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSP LGITDWN SVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLSEYGNLISACVLFIIDEVKRKRSM AEGKSTTGEGLDCGVLF GFGPGMTVETVVLRSVRVTA AAVAN
GN

> 2-PS variant= 4 Substitutions= 1: (L202F)
MGYSDDVEVIREAGRAQGLATIIAIGTATPPNCVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI VCSEITAIffFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTI LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSP LGITDWN SVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLSEYGNLISACVLFIIDEVKRKRSM AEGKSTTGEGLDCGVLF GFGPGMTVETVVLRSVRVTA AAVAN
GN

```

## FIG. 10 (continued)

```

> 2-PS variant= 5 Substitutions= 1: (L268F)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGT[VLRLAKDLAENNKGSRLIVCSEITAI[LFHGPENENHLD[SLVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTI[LPDTEKAMKLHLREGG[TFQLHRDVPLMVAKNIENAAEKALSP[IGITDWSVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKR[SM[AE[KGSTTGEGLDCGVLFEGFGP[GMTVETVVLRSVRVTA[AVAN
GN

> 2-PS variant= 6 Substitutions= 1: (L268M)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGT[VLRLAKDLAENNKGSRLIVCSEITAI[LFHGPENENHLD[SLVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTI[LPDTEKAMKLHLREGG[TFQLHRDVPLMVAKNIENAAEKALSP[IGITDWSVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKR[SM[AE[KGSTTGEGLDCGVLFEGFGP[GMTVETVVLRSVRVTA[AVAN
GN

> 2-PS variant= 7 Substitutions= 2: (L202F) (L268M)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGT[VLRLAKDLAENNKGSRLIVCSEITAI[LFHGPENENHLD[SLVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTI[LPDTEKAMKLHLREGG[TFQLHRDVPLMVAKNIENAAEKALSP[IGITDWSVFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKR[SM[AE[KGSTTGEGLDCGVLFEGFGP[GMTVETVVLRSVRVTA[AVAN
GN

```

## FIG. 10 (continued)

> 2-PS variant= 8 Substitutions= 2: (C 89S) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMSEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRASRHLSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDSGVLFEGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 9 Substitutions= 3: (C 89S) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMSEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRASRHLSEYGNLISAAVLFIIDEVRKRSMAGKSTTGEGLDSGVLFEGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 10 Substitutions= 3: (C 65S) (C 89S) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYFRVTKSEHMDLKEKFKRISEKTAIKKRYLALTED  
 YLQENPTMSEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRASRHLSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDSGVLFEGFGPGMTVETVILRSVRVTAAVAN  
 GN

## FIG. 10 (continued)

> 2-PS variant= 11 Substitutions= 4: (C 89S) (C195S) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYYFRVTKSEHMDLKEKFKRI[SEKTAIKKRYLALTED  
 YLQENPTM[SEFFMAPSLNARQDLVVTGVPMGLGKEAAVKAI[DEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLI[SEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKN[IEENAAEKALSP[LGITDWN[SVFWMVHPGGRAILDQVER  
 KLN[KEDKLRASRHLSEYGNLISAA[VLFIID[EVKRKR[SM[AE[KGKSTTGEGLD[SGVLFGFGPGMTVETV[VLRSVRVTA[AAVAN

GN

> 2-PS variant= 12 Substitutions= 5: (C 65S) (C 89S) (C195S) (C346A) (C372S) 18/150  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCVAQADYADYYFRVTKSEHMDLKEKFKRI[SEKTAIKKRYLALTED  
 YLQENPTM[SEFFMAPSLNARQDLVVTGVPMGLGKEAAVKAI[DEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLI[SEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKN[IEENAAEKALSP[LGITDWN[SVFWMVHPGGRAILDQVER  
 KLN[KEDKLRASRHLSEYGNLISAA[VLFIID[EVKRKR[SM[AE[KGKSTTGEGLD[SGVLFGFGPGMTVETV[VLRSVRVTA[AAVAN

GN

> 2-PS variant= 13 Substitutions= 8: (C 35S) (C 65S) (C 89S) (C195S) (M259L)  
 (L261N) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYYFRVTKSEHMDLKEKFKRI[SEKTAIKKRYLALTED  
 YLQENPTM[SEFFMAPSLNARQDLVVTGVPMGLGKEAAVKAI[DEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLI[SEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKA[KH]HLREGGLTFQLHRDVPLMVAKN[IEENAAEKALSP[LGITDWN[SVFWMVHPGGRAILDQVER  
 KLN[KEDKLRASRHLSEYGNLISAA[VLFIID[EVKRKR[SM[AE[KGKSTTGEGLD[SGVLFGFGPGMTVETV[VLRSVRVTA[AAVAN

GN

## FIG. 10 (continued)

> 2-PS variant= 14 Substitutions= 10: (C 35S) (C 65S) (C 89S) (T137F)  
 (C195S) (L202G) (M259L) (L261N) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEH[MVDLKEFKR][S]EKTAIKKRYLALTED  
 YLQENPTM[S]EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCT[F]AGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIV[S]SEITAI[C]FHGPNENHLDLSVAQALFGDGAAALIVSGPHLAVERP  
 IFEIVSTDQTI LPDTEKA[IK][HL]REGGLTFQLHRDVPLMVAKNIENAAEKALSP LGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRSRHLVSEYGNLISA[a]VLFIIDEVRKRSMAGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSVRTAAVAN  
 GN

> 2-PS variant= 15 Substitutions= 10: (C 35S) (C 65S) (C 89S) (C195S)  
 (I201F) (L202G) (M259L) (L261N) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEH[MVDLKEFKR][S]EKTAIKKRYLALTED  
 YLQENPTM[S]EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIV[S]SEITAI[f]FHGPNENHLDLSVAQALFGDGAAALIVSGPHLAVERP  
 IFEIVSTDQTI LPDTEKA[IK][HL]REGGLTFQLHRDVPLMVAKNIENAAEKALSP LGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRSRHLVSEYGNLISA[a]VLFIIDEVRKRSMAGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSVRTAAVAN  
 GN

> 2-PS variant= 16 Substitutions= 10: (C 35S) (C 65S) (C 89S) (T137F)  
 (C195S) (L202G) (M259L) (L261N) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEH[MVDLKEFKR][S]EKTAIKKRYLALTED  
 YLQENPTM[S]EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCT[F]AGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIV[S]SEITAI[g]FHGPNENHLDLSVAQALFGDGAAALIVSGPHLAVERP  
 IFEIVSTDQTI LPDTEKA[IK][HL]REGGLTFQLHRDVPLMVAKNIENAAEKALSP LGITDWNVSFWMVHPGGRAILDQVER  
 KLNKEDKLRSRHLVSEYGNLISA[a]VLFIIDEVRKRSMAGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSVRTAAVAN  
 GN

## FIG. 10 (continued)

```

> 2-PS variant= 17 Substitutions= 10: (C 35S) (C 65S) (C 89S) (C195S)
(L202G) (M259L) (L261N) (L268M) (C346A) (C372S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFRVTKSEHMDLKEFKRIS]EKTAIKKRYLALTED
YLQENPTM[EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI]THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIV[SEITAI]G[FHGPENENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKA[IK]HLREGG[TFQLHRDVP]LMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDDQVER
KLNLKEDKLRSRHLVSEYGNLISA[AVLFIIDEVRKR]SMAEGKSTTGEGLD[SGVLF]GFGPGMTVETVLR[SVR]VTAAVAN
GN

> 2-PS variant= 18 Substitutions= 3: (C 35S) (C 89S) (C372S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFRVTKSEHMDLKEFKRICEKTAI]KKRYLALTED
YLQENPTM[EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI]THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPENENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLREGGLTFQLHRDVP]LMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKR]SMAEGKSTTGEGLD[SGVLF]GFGPGMTVETVLR[SVR]VTAAVAN
GN

> 2-PS variant= 19 Substitutions= 4: (C 35S) (C 89S) (C346A) (C372S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFRVTKSEHMDLKEFKRICEKTAI]KKRYLALTED
YLQENPTM[EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI]THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPENENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLREGGLTFQLHRDVP]LMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDDQVER
KLNLKEDKLRSRHLVSEYGNLISA[AVLFIIDEVRKR]SMAEGKSTTGEGLD[SGVLF]GFGPGMTVETVLR[SVR]VTAAVAN
GN

```

## FIG. 10 (continued)

> 2-PS variant= 20 Substitutions= 4: (C 35S) (C 65S) (C 89S) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRI[S]EKTAIKKRYLALTED  
 YLQENPTM[S]EFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>21</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLVSEYGNLISACVLFIIDEVKRKRMAEGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSRVTAAVAN  
 GN

> 2-PS variant= 21 Substitutions= 5: (C 35S) (C 89S) (C195S) (C346A) (C372S) 21/150  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTM[S]EFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>21</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIV[S]SEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLVSEYGNLISAA[a]VLFFIIDEVKRKRMAEGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSRVTAAVAN  
 GN

> 2-PS variant= 22 Substitutions= 6: (C 35S) (C 65S) (C 89S) (C195S) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRI[S]EKTAIKKRYLALTED  
 YLQENPTM[S]EFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>21</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIV[S]SEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLVSEYGNLISAA[a]VLFFIIDEVKRKRMAEGKSTTGEGLD[S]GVLFGFGPGMTVETVLRVSRVTAAVAN  
 GN

22/150

**FIG. 10 (continued)**

```

> 2-PS variant= 23 Substitutions= 2: (C 35S) (T137F)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNsVAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKKAIDEWGLPKSKI THLIFCTHAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP

IFEIVSTDQTIIPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVKRKRMAEGKSTTGEGLDCGVLFEGFGPGMTVETVILRSVRVTAAVAN
GN

> 2-PS variant= 24 Substitutions= 2: (C 35S) (T137M)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNsVAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKKAIDEWGLPKSKI THLIFCTHAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIIPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVKRKRMAEGKSTTGEGLDCGVLFEGFGPGMTVETVILRSVRVTAAVAN
GN

> 2-PS variant= 25 Substitutions= 2: (C 35S) (M259F)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNsVAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKKAIDEWGLPKSKI THLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIIPDTEKA[KKLH]LHREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVKRKRMAEGKSTTGEGLDCGVLFEGFGPGMTVETVILRSVRVTAAVAN
GN

```

23/150

**FIG. 10 (continued)**

> 2-PS variant= 26 Substitutions= 2: (C 35S) (T137L)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[**S**]VAQADYADYYFVRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCT[**I**]AGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>2</sup>LPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sup>3</sup>LGITD<sup>4</sup>WNSVFWMVHPGGRAILDQ<sup>5</sup>VER  
 KLN<sup>6</sup>LKEDKLRASRHLVSEYGNLISACVLFIIDEVKRKRSMAEGKSTTGEGLDCGVLF<sup>7</sup>GF<sup>8</sup>PGMTVETV<sup>9</sup>VLRSVRVTAAVAN

GN

> 2-PS variant= 27 Substitutions= 2: (C 35S) (I343F)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[**S**]VAQADYADYYFVRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>2</sup>LPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sup>3</sup>LGITD<sup>4</sup>WNSVFWMVHPGGRAILDQ<sup>5</sup>VER  
 KLN<sup>6</sup>LKEDKLRASRHLVSEYGNL[**I**]SACVLFIIDEVKRKRSMAEGKSTTGEGLDCGVLF<sup>7</sup>GF<sup>8</sup>PGMTVETV<sup>9</sup>VLRSVRVTAAVAN

GN

> 2-PS variant= 28 Substitutions= 2: (C 35S) (I343M)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[**S**]VAQADYADYYFVRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>2</sup>LPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sup>3</sup>LGITD<sup>4</sup>WNSVFWMVHPGGRAILDQ<sup>5</sup>VER  
 KLN<sup>6</sup>LKEDKLRASRHLVSEYGNL[**I**]SACVLFIIDEVKRKRSMAEGKSTTGEGLDCGVLF<sup>7</sup>GF<sup>8</sup>PGMTVETV<sup>9</sup>VLRSVRVTAAVAN

GN

## FIG. 10 (continued)

> 2-PS variant= 29 Substitutions= 2: (C 35S) (I201F)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAI<sup>2</sup>[f]LFHGPNNHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>3</sup>LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sup>4</sup>LGITDWN<sup>5</sup>SVFWMVHPGGRAILDQVER  
 KLN<sup>6</sup>LKEDKLRASRHLSEYGNLISACVLFIIDEV<sup>7</sup>RKR<sup>8</sup>SMAEGKSTTGEGLDCGVLF<sup>9</sup>FGPGMTVETV<sup>10</sup>LV<sup>11</sup>RSVRVTAAVAN

GN

> 2-PS variant= 30 Substitutions= 2: (C 35S) (C169A)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQG[a]AAGGTVLRLAKDLAENNKGSRVLIVCSEITAI<sup>2</sup>LFHGPNNHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>3</sup>LPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sup>4</sup>LGITDWN<sup>5</sup>SVFWMVHPGGRAILDQVER  
 KLN<sup>6</sup>LKEDKLRASRHLSEYGNLISACVLFIIDEV<sup>7</sup>RKR<sup>8</sup>SMAEGKSTTGEGLDCGVLF<sup>9</sup>FGPGMTVETV<sup>10</sup>LV<sup>11</sup>RSVRVTAAVAN

GN

> 2-PS variant= 31 Substitutions= 2: (C 35S) (L268M)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI<sup>1</sup>THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAI<sup>2</sup>LFHGPNNHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI<sup>3</sup>LPDTEKAMKHLHREGG[m]TFQLHRDVPLMVAKNIEAAEKALSP<sup>4</sup>LGITDWN<sup>5</sup>SVFWMVHPGGRAILDQVER  
 KLN<sup>6</sup>LKEDKLRASRHLSEYGNLISACVLFIIDEV<sup>7</sup>RKR<sup>8</sup>SMAEGKSTTGEGLDCGVLF<sup>9</sup>FGPGMTVETV<sup>10</sup>LV<sup>11</sup>RSVRVTAAVAN

GN

25/150

## FIG. 10 (continued)

> 2-PS variant= 32 Substitutions= 2: (C 35S) (L202F)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAT[L]FHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLSEYGNLISACVLFIIDEVRKRSMAGCKSTTGEGLDCGVLF[FGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 33 Substitutions= 2: (C 35S) (I201M)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAT[L]FHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLSEYGNLISACVLFIIDEVRKRSMAGCKSTTGEGLDCGVLF[FGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 34 Substitutions= 2: (C 35S) (I201L)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAT[L]FHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITDWNVSFFWMVHPGGRAILDQVER  
 KLNLKEDKLRASRHLSEYGNLISACVLFIIDEVRKRSMAGCKSTTGEGLDCGVLF[FGFGPGMTVETVILRSVRVTAAVAN  
 GN

26/150

## FIG. 10 (continued)

> 2-PS variant= 35 Substitutions= 2: (C 35S) (L202M)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSINARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPCADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAI[H]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTILPDTEKAMKHLREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDDQVER  
 KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 36 Substitutions= 4: (C 35S) (L202G) (M259L) (L261N)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSINARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPCADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAI[G]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTILPDTEKA[K][N]HLREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDDQVER  
 KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 37 Substitutions= 9: (C 35S) (C 65S) (C 89S) (C195S) (L202G)  
 (M259L) (L261N) (C346A) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYFVRVTKSEHMVDLKEKFKRI[S]EKTAKKRYLALTED  
 YLQENPTM[S]EFMAPSINARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPCADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIV[S]SEITAI[G]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTILPDTEKA[K][N]HLREGGLTFQLHRDVPLMVAKNIEAAEKALSPLGITDWNVSFWMVHPGGRAILDDQVER  
 KLNLKEDKLRSRHLVSEYGNLISa[a]VLFIIDEVRKRSMAGKSTTGEGLD[S]GVLFGFGPGMTVETVILRSVRVTAAVAN  
 GN

27/150

## FIG. 10 (continued)

```

> 2-PS variant= 38 Substitutions= 2: (C 35S) (L202G)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEFKRICEKTAIKKRYLALTED
YLQENPTMCEFMPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTIVRLAKDLAENNKGSRLIVCSEITAI[C]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGLGTDWNSVFWMVHPGGRAILDQVER
KLNKEDKLRSRHLSEYGNLISACVLFIDEVRKRSMAGCKSTTGEGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 39 Substitutions= 3: (C 35S) (M259L) (L261N)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEFKRICEKTAIKKRYLALTED
YLQENPTMCEFMPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTIVRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTILPDTEKA[KL]HLREGGLTFQLHRDVPLMVAKNIENAAEKALSPGLGTDWNSVFWMVHPGGRAILDQVER
KLNKEDKLRSRHLSEYGNLISACVLFIDEVRKRSMAGCKSTTGEGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 40 Substitutions= 7: (C 35S) (C 65S) (C 89S) (C195S) (L202G)
(C346A) (C372S)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEFKRI[S]EKTAIKKRYLALTED
YLQENPTM[S]EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTIVRLAKDLAENNKGSRLIV[S]SEITAI[C]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPGLGTDWNSVFWMVHPGGRAILDQVER
KLNKEDKLRSRHLSEYGNLISA[a]VLFIIDEVRKRSMAGCKSTTGEGLD[S]GVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

```

28/150

## FIG. 10 (continued)

> 2-PS variant= 41 Substitutions= 1: (C 35S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGT<sub>VL</sub>RLAKDLAENNKGSRVLI<sub>VC</sub>SEITAI[L]FHGPNENHLD<sub>SL</sub>VAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQ<sub>TIL</sub>PDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIEAAEKALSP<sub>L</sub>GITD<sub>WNS</sub>VFWMVHPGGRAILDQ<sub>VER</sub>  
 KNLKEDKLRSRHLSEYGNLISACVLFIIDEVRKR<sub>SM</sub>AECKSTTGEGLDCGVLF<sub>GF</sub>GP<sub>GM</sub>T<sub>VE</sub>TVVLR<sub>SV</sub>RVTA<sub>AVAN</sub>  
 GN

> 2-PS variant= 42 Substitutions= 4: (C 35S) (T137L) (L202F) (L268M)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCT[L]AGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGT<sub>VL</sub>RLAKDLAENNKGSRVLI<sub>VC</sub>SEITAI[L]FHGPNENHLD<sub>SL</sub>VAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQ<sub>TIL</sub>PDTEKAMKLHLREGG[M]TFQLHRDVPLMVAKNIEAAEKALSP<sub>L</sub>GITD<sub>WNS</sub>VFWMVHPGGRAILDQ<sub>VER</sub>  
 KNLKEDKLRSRHLSEYGNLISACVLFIIDEVRKR<sub>SM</sub>AECKSTTGEGLDCGVLF<sub>GF</sub>GP<sub>GM</sub>T<sub>VE</sub>TVVLR<sub>SV</sub>RVTA<sub>AVAN</sub>  
 GN

> 2-PS variant= 43 Substitutions= 5: (C 35S) (T137M) (L202F) (L268M) (I343F)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCT[M]AGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGT<sub>VL</sub>RLAKDLAENNKGSRVLI<sub>VC</sub>SEITAI[L]FHGPNENHLD<sub>SL</sub>VAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQ<sub>TIL</sub>PDTEKAMKLHLREGG[M]TFQLHRDVPLMVAKNIEAAEKALSP<sub>L</sub>GITD<sub>WNS</sub>VFWMVHPGGRAILDQ<sub>VER</sub>  
 KNLKEDKLRSRHLSEYGNL[F]SACVLFIIDEVRKR<sub>SM</sub>AECKSTTGEGLDCGVLF<sub>GF</sub>GP<sub>GM</sub>T<sub>VE</sub>TVVLR<sub>SV</sub>RVTA<sub>AVAN</sub>  
 GN

## FIG. 10 (continued)

> 2-PS variant= 44 Substitutions= 4: (C 35S) (L202F) (L268M) (I343F)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAI[D]EWGLPKSKI[T]HLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAI[F]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI[L]PDTEKAMKLHLREGG[T]TFQLHRDVP[L]MVAKN[I]ENAAEKALSP[L]GITD[N]SVFWMVHPGGR[A]ILDQ[V]ER  
 KLN[L]KEDKLRASR[H]VLSEYGN[I]F[S]ACVLFII[D]EVRKR[S]MAEGKSTTGEGLDCGVLF[G]FGPGMTVETV[L]RSVRVTAAVAN  
 GN

> 2-PS variant= 45 Substitutions= 2: (C 35S) (C372S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAI[D]EWGLPKSKI[T]HLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAI[L]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI[L]PDTEKAMKLHLREGGLTFQLHRDVP[L]MVAKN[I]ENAAEKALSP[L]GITD[N]SVFWMVHPGGR[A]ILDQ[V]ER  
 KLN[L]KEDKLRASR[H]VLSEYGN[L]ISACVLFII[D]EVRKR[S]MAEGKSTTGEGLD[S]GVLF[G]FGPGMTVETV[L]RSVRVTAAVAN  
 GN

> 2-PS variant= 46 Substitutions= 2: (C 35S) (C372A)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAI[D]EWGLPKSKI[T]HLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAI[L]FHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTI[L]PDTEKAMKLHLREGGLTFQLHRDVP[L]MVAKN[I]ENAAEKALSP[L]GITD[N]SVFWMVHPGGR[A]ILDQ[V]ER  
 KLN[L]KEDKLRASR[H]VLSEYGN[L]ISACVLFII[D]EVRKR[S]MAEGKSTTGEGLD[A]GVLF[G]FGPGMTVETV[L]RSVRVTAAVAN  
 GN

## FIG. 10 (continued)

> 2-PS variant= 47 Substitutions= 2: (C 35S) (C346S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[s]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIIPDTEKAMKHLHREGGLTFQLHRDVPMLVAKNIENAAEKALSPLGITDWNVSFFWMVHPGGRAILDQVER  
 KLNKEDKLRASRVHVLSEYGNLISA[s]VLFIIDEVRKRSMAGKSTTGEGLDCCGVLFGGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 48 Substitutions= 2: (C 35S) (C346A)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[s]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIIPDTEKAMKHLHREGGLTFQLHRDVPMLVAKNIENAAEKALSPLGITDWNVSFFWMVHPGGRAILDQVER  
 KLNKEDKLRASRVHVLSEYGNLISA[a]VLFIIDEVRKRSMAGKSTTGEGLDCCGVLFGGPGMTVETVILRSVRVTAAVAN  
 GN

> 2-PS variant= 49 Substitutions= 2: (C 35S) (C195S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[s]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI[s]SEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP  
 IFEIVSTDQTIIPDTEKAMKHLHREGGLTFQLHRDVPMLVAKNIENAAEKALSPLGITDWNVSFFWMVHPGGRAILDQVER  
 KLNKEDKLRASRVHVLSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCCGVLFGGPGMTVETVILRSVRVTAAVAN  
 GN

31/150

## FIG. 10 (continued)

> 2-PS variant= 50 Substitutions= 2: (C 35S) (C195A)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEH[MVDLKEFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLI]FCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI[Va]SEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKAMK]HLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDQVER  
 KLN[KEDK]LRASRHLSEYGNLISACVLFIIDEVKR[RSMAEGKSTTGEGLDCGVLFGFGPGMTVETV]LRSVRVTAAVAN  
 GN

> 2-PS variant= 51 Substitutions= 2: (C 35S) (C135S)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEH[MVDLKEFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLI]FCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI[VCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKAMK]HLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDQVER  
 KLN[KEDK]LRASRHLSEYGNLISACVLFIIDEVKR[RSMAEGKSTTGEGLDCGVLFGFGPGMTVETV]LRSVRVTAAVAN  
 GN

> 2-PS variant= 52 Substitutions= 2: (C 35S) (C135A)  
 MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEH[MVDLKEFKRICEKTAIKKRYLALTED  
 YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI[THLI]FCTTAGVDMPGADYQLVKLLGLSPSVK  
 RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLI[VCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP  
 IFEIVSTDQTI[LPDTEKAMK]HLREGGLTFQLHRDVPLMVAKNIENAAEKALSP[LGITD]WNSVFWMVHPGGRAILDQVER  
 KLN[KEDK]LRASRHLSEYGNLISACVLFIIDEVKR[RSMAEGKSTTGEGLDCGVLFGFGPGMTVETV]LRSVRVTAAVAN  
 GN

32/150

## FIG. 10 (continued)

```

> 2-PS variant= 53 Substitutions= 2: (C 35S) (C 89S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTM[EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER
KLNKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 54 Substitutions= 2: (C 35S) (C 89A)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYYFRVTKSEHVMVDLKEKFKRICEKTAIKKRYLALTED
YLQENPTM[EFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER
KLNKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 55 Substitutions= 2: (C 35S) (C 65S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYYFRVTKSEHVMVDLKEKFKRI[SEKTAIKKRYLALTED
YLQENPTMCEEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMVAKNIENAAEKALSPGITDWNVSFWMVHPGGRAILDQVER
KLNKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

```

33/150

**FIG. 10 (continued)**

```

> 2-PS variant= 56 Substitutions= 2: (C 35S) (C 65A)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRI[a]EKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 57 Substitutions= 2: (C 35S) (C169S)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQG[S]AAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGNLISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 58 Substitutions= 2: (C 35S) (N341H)
MGSYSSDDVEVIREAGRAQGLATILAIGTATPPN[S]VAQADYADYYFRVTKSEHMDLKEKFKRICEKTAIKKRYLALTED
YLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKIITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGGTVLRLAKDLAENNKGSRLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNLKEDKLRSRHLVSEYGN[h]LISACVLFIIDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVVLRSVRVTAAVAN
GN

```

34/150

## FIG. 10 (continued)

```

> 2-PS variant= 59 Substitutions= 2: (C 35S) (H308Q)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFVRVTKSEHMDLKEFKRICEKTAIKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGGLTFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMV[PGGRAILDQVER
KLNKEDKLRSRHLVSEYGNLISACVLFIIDEVKRKRSMAEGKSTTGGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 60 Substitutions= 3: (C 35S) (L202F) (L268F)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFVRVTKSEHMDLKEFKRICEKTAIKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLIVCSEITAI[FHGPENENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGG[TFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNKEDKLRSRHLVSEYGNLISACVLFIIDEVKRKRSMAEGKSTTGGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

> 2-PS variant= 61 Substitutions= 3: (C 35S) (L268M) (I343M)
MGYSDDVEVIREAGRAQGLATILAIGTATPPN[VAQADYADYFVRVTKSEHMDLKEFKRICEKTAIKRYLALTED
YLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSPSVK
RYMLYQQGCAAGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENHLDLSVAQALFGDGAALIVGSGPHLAVERP
IFEIVSTDQTIILPDTEKAMKHLHREGG[TFQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER
KLNKEDKLRSRHLVSEYGNL[ISACVLFIIDEVKRKRSMAEGKSTTGGLDCGVLFGEFGPGMTVETVVLRSVRVTAAVAN
GN

```

## FIG. 10 (continued)

> 2-PS variant= 62 Substitutions= 3: (C 35S) (L202F) (L268M)  
MGSYSSDDVEVIREAGRAQGLATIIAIGTATPPNSVAQADYADYYFRVTKSEHMYDLKEKFKRICEKTAIKKRYLALTED  
YLQENPTMCEFFMAPSLNARQDLVVTGVPMGLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMFGADYQLVKLLGLSPSVK  
RYMLYQQGCAAGTVLRLAKDLAENNKGSRLIVCSEITAIFFHGPENENHLDLSVAQALFQDGAALIVGSGPHLAVERP  
IFEIVSTDQTIILPDTEKAMKHLHREGGTFQQLHRDVPLMVAKNIENAAEKALSPLGITDWNVSFWMVHPGGRAILDQVER  
KLNKEDKLRASRHVLSEYGNLISACVLFIIDEVKRKRSMAEGKSTTGEGLDCGVLFQFGPGMTVETVVLRSVRVTAAVAN  
GN



## FIG. 11 (continued)

> Seq 22 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 23 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 24 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 25 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 26 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 27 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 28 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 29 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 30 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 31 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 32 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 33 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 34 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 35 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 36 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 37 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 38 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 39 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 40 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 41 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 42 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 43 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 44 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI  
> Seq 45 MGSYSSDDVEVIREAGRAQGLATILAI GTATPPNsVAQADYADYFRVTKSEHMVDLKEKFKRISekTAI

FIG. 11 (continued)

|       |    |                             |                       |                       |
|-------|----|-----------------------------|-----------------------|-----------------------|
| > Seq | 46 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 47 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 48 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 49 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 50 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 51 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 52 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 53 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 54 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 55 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 56 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 57 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 58 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 59 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 60 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 61 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |
| > Seq | 62 | MGSYSSDDVEVIREAGRAQGLATILAI | GTATPPNsVAQADYADYFRVT | KSEHMVDLKEKFKRICEKTAI |

|       |    |                        |                              |                      |     |
|-------|----|------------------------|------------------------------|----------------------|-----|
| > Seq | 71 | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV | 140 |
| > Seq | 1  | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV |     |
| > Seq | 2  | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV |     |
| > Seq | 3  | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV |     |
| > Seq | 4  | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV |     |
| > Seq | 5  | KKRYLALTEDYLQENPTMCEFM | APSLNARQDLVVTGVPMLGKEAAVKAID | EWGLPKSKITHLIFCTTAGV |     |

39/150

## FIG. 11 (continued)

> Seq 6 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 7 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 8 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 9 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 10 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 11 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 12 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 13 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 14 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 15 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 16 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 17 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 18 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 19 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 20 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 21 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 22 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 23 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 24 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 25 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 26 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 27 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 28 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV  
> Seq 29 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKI THLI FCTTAGV

## FIG. 11 (continued)

> Seq 30 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 31 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 32 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 33 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 34 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 35 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 36 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 37 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 38 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 39 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 40 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 41 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 42 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 43 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 44 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 45 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 46 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 47 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 48 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 49 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 50 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 51 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFsTTAGV  
 > Seq 52 KKRYLALTEDYLQENPTMCEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFaTTAGV  
 > Seq 53 KKRYLALTEDYLQENPTMsEFFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV

FIG. 11 (continued)

> Seq 54 KKRYLALTEDYLQENPTMaEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 55 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 56 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 57 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 58 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 59 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 60 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 61 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV  
 > Seq 62 KKRYLALTEDYLQENPTMCEFMAPSLNARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGV

141 . . . . . 210  
 > Seq 1 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 2 DMPGADYQLVKLLGLSPSVKRYMLYQQGaAAGTTLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 3 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 4 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 5 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 6 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 7 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 8 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 9 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 10 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHGPNENH  
 > Seq 11 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVsSEITAILFHGPNENH  
 > Seq 12 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVsSEITAILFHGPNENH  
 > Seq 13 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVsSEITAILFHGPNENH

## FIG. 11 (continued)

> Seq 14 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFhGPNENH  
 > Seq 15 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFhGPNENH  
 > Seq 16 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFhGPNENH  
 > Seq 17 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFhGPNENH  
 > Seq 18 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 19 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 20 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 21 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAILfHGPENH  
 > Seq 22 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAILfHGPENH  
 > Seq 23 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 24 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 25 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 26 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 27 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 28 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 29 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAIflfHGPENH  
 > Seq 30 DMPGADYQLVKLLGLSPSVKRYMLYQQGaAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 31 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 32 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAIffhGPNENH  
 > Seq 33 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAmLfHGPENH  
 > Seq 34 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAILfHGPENH  
 > Seq 35 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAImlfHGPENH  
 > Seq 36 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvcSEITAIgFhGPNENH  
 > Seq 37 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFhGPNENH

## FIG. 11 (continued)

> Seq 38 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAIgFHHGPNENH  
 > Seq 39 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 40 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAIgFHHGPNENH  
 > Seq 41 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 42 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAIffHHGPNENH  
 > Seq 43 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAIffHHGPNENH  
 > Seq 44 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAIffHHGPNENH  
 > Seq 45 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 46 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 47 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 48 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 49 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvsSEITAILFHHGPNENH  
 > Seq 50 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIvaSEITAILFHHGPNENH  
 > Seq 51 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 52 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 53 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 54 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 55 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 56 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 57 DMPGADYQLVKLLGLSPSVKRYMLYQQGsAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 58 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 59 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH  
 > Seq 60 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAIffHHGPNENH  
 > Seq 61 DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGGTVLRLAKDLAENNKGSRVLIVCSEITAILFHHGPNENH

## FIG. 11 (continued)

|    |     |  |     |                                   |                                       |     |
|----|-----|--|-----|-----------------------------------|---------------------------------------|-----|
| >> | Seq |  | 62  | DMPGADYQLVKLLGLSPSVKRYMLYQQGCAAGT | VLRRLAKDLAENNKGSRVLIVCSEITAIFFHGPNEHH |     |
| >> | Seq |  | 211 | .                                 | .                                     | 280 |
| >> | Seq |  | 1   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 2   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 3   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 4   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 5   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 6   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 7   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 8   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 9   | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 10  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 11  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 12  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 13  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAlKnHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 14  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAlKnHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 15  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAlKnHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 16  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAlKnHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 17  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAlKnHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 18  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 19  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 20  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |
| >> | Seq |  | 21  | LDSLVAQALFGDGAAALIVSGPHLAVERPIFEI | IVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP    | LMV |

## FIG. 11 (continued)

> Seq 22 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 23 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 24 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 25 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAFKLHLREGGLTFQLHRDVPLMV  
 > Seq 26 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 27 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 28 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 29 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 30 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 31 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGmTFQLHRDVPLMV  
 > Seq 32 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 33 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 34 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 35 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 36 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAlKnHLREGGLTFQLHRDVPLMV  
 > Seq 37 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAlKnHLREGGLTFQLHRDVPLMV  
 > Seq 38 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 39 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAlKnHLREGGLTFQLHRDVPLMV  
 > Seq 40 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 41 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV  
 > Seq 42 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGmTFQLHRDVPLMV  
 > Seq 43 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGmTFQLHRDVPLMV  
 > Seq 44 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGmTFQLHRDVPLMV  
 > Seq 45 LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQTIILPDTEKAMKLHLREGGLTFQLHRDVPLMV

## FIG. 11 (continued)

|    |     |    |                                                                          |
|----|-----|----|--------------------------------------------------------------------------|
| >> | Seq | 46 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 47 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 48 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 49 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 50 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 51 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 52 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 53 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 54 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 55 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 56 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 57 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 58 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 59 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 60 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 61 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |
| >> | Seq | 62 | LDSLVAQALFGDGAAALIVGSGPHLAVERPIFEIVSTDQQTILPDTEKAMKHLHREGGLTFQLHRDVPPLMV |

|     |     |   |                |     |              |       |           |         |          |          |     |
|-----|-----|---|----------------|-----|--------------|-------|-----------|---------|----------|----------|-----|
| 281 | .   | . | .              | .   | .            | .     | 350       |         |          |          |     |
| >   | Seq | 1 | AKNIENAAEKALSP | LGI | TDWNSVFWMVHP | GGRAI | LDQVERKLN | LKEDKLR | SRHVLSEY | GNLISACV | LFI |
| >   | Seq | 2 | AKNIENAAEKALSP | LGI | TDWNSVFWMVHP | GGRAI | LDQVERKLN | LKEDKLR | SRHVLSEY | GNLISACV | LFI |
| >   | Seq | 3 | AKNIENAAEKALSP | LGI | TDWNSVFWMVHP | GGRAI | LDQVERKLN | LKEDKLR | SRHVLSEY | GNLISACV | LFI |
| >   | Seq | 4 | AKNIENAAEKALSP | LGI | TDWNSVFWMVHP | GGRAI | LDQVERKLN | LKEDKLR | SRHVLSEY | GNLISACV | LFI |
| >   | Seq | 5 | AKNIENAAEKALSP | LGI | TDWNSVFWMVHP | GGRAI | LDQVERKLN | LKEDKLR | SRHVLSEY | GNLISACV | LFI |

FIG. 11 (continued)

> Seq 6 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 7 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 8 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 9 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 10 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 11 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 12 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 13 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 14 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 15 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 16 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 17 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 18 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 19 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 20 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 21 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 22 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISaAVLFI  
 > Seq 23 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 24 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 25 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 26 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI  
 > Seq 27 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLfSACVLFI  
 > Seq 28 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLmSACVLFI  
 > Seq 29 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKLNKEDKLRASRHVLSSEYGNLISACVLFI

## FIG. 11 (continued)

> Seq 30 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 31 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 32 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 33 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 34 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 35 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 36 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 37 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISAAVLFI  
> Seq 38 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 39 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 40 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISAAVLFI  
> Seq 41 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 42 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 43 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 44 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 45 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 46 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 47 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISASVLFI  
> Seq 48 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISAAVLFI  
> Seq 49 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 50 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 51 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 52 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI  
> Seq 53 AKNIENAAEKALSPGLGTTDWSVFWMVHPGGRAILDQVERKINLKEDKLRSRHHVLSYGNLISACVLFI

FIG. 11 (continued)

> Seq 54 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 55 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 56 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 57 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 58 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 59 AKNIENAAEKALSPGLGTTDWNVSFWMVqPPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 60 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF  
 > Seq 61 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLmSACVLF  
 > Seq 62 AKNIENAAEKALSPGLGTTDWNVSFWMVHPGGRAILDQVERKLNKEDKLRSRHLSEYGNLISACVLF

351 . . . 402  
 > Seq 1 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 2 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 3 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 4 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 5 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 6 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 7 IDEVRKRSMAGKSTTGEGLDCGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 8 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 9 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 10 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 11 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 12 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN  
 > Seq 13 IDEVRKRSMAGKSTTGEGLDsGVLFGFGPGMTVETVLRSVRVTAAVANGN

## FIG. 11 (continued)

> Seq 14 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 15 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 16 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 17 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 18 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 19 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 20 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 21 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 22 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 23 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 24 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 25 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 26 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 27 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 28 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 29 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 30 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 31 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 32 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 33 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 34 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 35 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 36 IDEVRKRSM AEGKSTT GEGLDcGVLF GFGPGMTVETVVLRSVRVTAAVANGN  
> Seq 37 IDEVRKRSM AEGKSTT GEGLDsGVLF GFGPGMTVETVVLRSVRVTAAVANGN

## FIG. 11 (continued)

> Seq 38 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 39 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 40 IDEVRKRSM AEGKSTT GEGLD sGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 41 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 42 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 43 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 44 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 45 IDEVRKRSM AEGKSTT GEGLD sGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 46 IDEVRKRSM AEGKSTT GEGLD aGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 47 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 48 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 49 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 50 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 51 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 52 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 53 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 54 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 55 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 56 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 57 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 58 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 59 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 60 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 61 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN  
> Seq 62 IDEVRKRSM AEGKSTT GEGLD CGVLF GFGPGMTVETV LRSVRVTA AVANGN

52/150

## FIG. 12

```

> 2-PS variant= 2 Substitutions= 1: (C169A)
ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCCGGACAGGTTTAGCCACGATTCTTGCCCAT
TGGCACTGCTACTCCTCCCAAATTGCGTCGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA
TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAGAAACGATACCTAGCCCTCACCGAAGAC
TATCTGCAAGAACCCCAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTCAACCGCGGT
CCCAATGCTTGGCAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT
TCTGCACCAACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA
CGCTATATGTTGTACCAACAGGAGGCTGCCGCCGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAACAAACAA
GGCTCACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC
TCGTCCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCAGTTGGCCGTAGAACGGCCA
ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG
GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC
CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCAACCCAGTGCTGAGCCATATTGGACCAAGGTGGAGCGA
AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT
GTTGTTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCCGGAAGGAAGATACAACCGGTGAAGGTTTGGATTGCGGTG
TTTGTGTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT
GGAAACTGA

```

```

> 2-PS variant= 3 Substitutions= 1: (C 35A)
ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCCGGACAGGTTTAGCCACGATTCTTGCCCAT
TGGCACTGCTACTCCTCCCAAATGCCGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA
TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAGAAACGATACCTAGCCCTCACCGAAGAC
TATCTGCAAGAACCCCAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTCAACCGCGGT
CCCAATGCTTGGCAAGAACCCGAGTCAAGGCCATTGATGAATGGGCACTACCAAAATCCAAAGATCACCCACCTCATCT

```

53/150

## FIG. 12 (continued)

TCTGCACCAACGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAGGACCTTGTCTGAAACAAACAA  
 GGGCTACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCAAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGAACTCAGTTTCTGGATGGTGCAACCCAGTGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTCACGAGGTGAGGAAGAGATCTATGCGCGAAGGAAGAGTACAACCGGTGAAGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 4 Substitutions= 1: (L202F)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTTCGTGAGGCCGGACGGGCACAAGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATATTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAACCAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTACCCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAGGACCTTGTCTGAAACAAACAA  
 GGGCTACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTCCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCGGAGAAAGCGTTGTCTC

## FIG. 12 (continued)

CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGTGGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCCGGAAGGAAGAGTACAACCGGTGAAGGTTTGATTTGCCGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCCGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 5 Substitutions= 1: (L268F)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCGGACGCGCAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGCCGT  
 CCCAATGCTTGGCAAGAACCCGCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCCGGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCACGAGTCCTTATCGTCTGTCTCCGAGATCACTGCTATCTTATTCCATGGACCCAAATGAGAACCACTTGACTCAC  
 TCGTC GCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GttcACGTTTCAGTTGCATAGAGATGACCCCTTGATGGTCGAAAGAACATAGAGAACGACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTCGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCCGGAAGGAAGAGTACAACCGGTGAAGGTTTGATTTGCCGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTGTGTTCTTCGTAGCGTCCGCTTACTGCTGCCGTTGCCAAT  
 GGAAACTGA

55/150

## FIG. 12 (continued)

> 2-PS variant= 6 Substitutions= 1: (L268M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGGTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCAITCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAACAAACAA  
 GGGCTCACGAGTCCTTATTCGTCTGCTCCGAGATCACTGCTATCTTATTTCCATGGACCCAAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 gatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTCGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGATGACTGTGAGACTGTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 7 Substitutions= 2: (L202F) (L268M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCAITCT

56/150

## FIG. 12 (continued)

TCTGCACCAACCGCTGCGTTGACATGCCCCGGTCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAAACAACAA  
 GGGCTCACGAGTCTTATCGTCTGCTCCGAGATCACCTGCTATCcttccatggaacccaatgagaaccacaccttgactcac  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCATTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 datgACGTTTCAGTTGCAATAGAGATGTACCCCTTGATGGTCGCAAGAACAATAGAGAACGACGAGAAAGCGTTGTCTC  
 CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCCGTTGAAGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCCGGTTACTGCTGCGGTTGCCCAAT  
 GGAAACTGA

> 2-PS variant= 8 Substitutions= 2: (C 89S) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCCGACGGCACAAAGGTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACCGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGctctGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGGT  
 CCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCGCACAGTCTCCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCAATAGAGATGTACCCCTTGATGGTCTGCAAGAACATAGAGAACGACGAGAAAGCGTTGTCTC

57/150

## FIG. 12 (continued)

CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGAAGGGAAGAGTACAACCGGTGAAGTTTGGAT**tcg**GGTG  
 TTTTGTGTTGGATTTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 9 Substitutions= 3: (C 89S) (C346A) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TCGTTGATCTTAAGAGAGAAATTAAACCGCATTTGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAACCCAAAT**gct**GAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGTGATGGTCCGCAAGAACATAGAGAACCGCAGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCT**gct**GT  
 GTTGTTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGAAGGGAAGAGTACAACCGGTGAAGTTTGGAT**tcg**GGTG  
 TTTTGTGTTGGATTTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

58/150

## FIG. 12 (continued)

> 2-PS variant= 10 Substitutions= 3: (C 65S) (C 89S) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGGTTAGCCACGATCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCAT~~ttct~~GAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 TATCTGCAAGAGAACCCAAACAATG~~ctct~~GAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGACCTTGCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCCGGACACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGACCCCTTGATGGTCGAAAGACATAGAGAACGCGAGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGCGAAGCAAGAGTACAAACCGGTGAAGGTTTGGAT~~tcct~~GGTG  
 TTTTGTGTGGATTTGACCGGGTATGACTGTTGAGACTGTTGTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 11 Substitutions= 4: (C 89S) (C195S) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGGTTTAGCCACGATCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATG~~ctct~~GAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAACCGCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

59/150

## FIG. 12 (continued)

TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCTCCGGCTAGCCAAGGACCTTGTCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTC[tcc]FCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTCATTTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGTCGCAAGAACAATAGAGAACGACGAGAAAGCGTTGTCTC  
 CACTAGGGATACTGATTGGAACTCAGTTTCTGGATGGTCGACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCT[gct]GT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGGGAACGGAAGAGTACAACCCGTGAAGGTTTGGAT[tcc]GGTG  
 TTTTGTTTGGATTTGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 12 Substitutions= 5: (C 65S) (C 89S) (C195S) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCCGGACGGCCACAAGGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAACGCGAT[tct]GAGAAACAGCGGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCAACAATG[tct]GAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAAGAAGCCGCACTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCAAGTCTCCGGCTAGCCAAGGACCTTGTCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTC[tcc]FCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGTCGCAAGAACAATAGAGAACGACGAGAAAGCGTTGTCTC

60/150

## FIG. 12 (continued)

CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTgctGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGATtccGGTG  
 TTTTGTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 13 Substitutions= 8: (C 35S) (C 65S) (C 89S) (C195S)

(M259L) (L261N) (C346A) (C372S)

ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCTGTGAGGCCGGACGGCCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTAAACGCATtctGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATgctGAGTTCATGGCTCCATCCTTAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAAAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCGGCTAGCCAAGGACCTTGCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTctccTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCAGTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAAGCActgAAQaaCAGTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACCGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAACCTGATTAGCGCTgctGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGATtccGGTG  
 TTTTGTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

61/150

## FIG. 12 (continued)

> 2-PS variant= 14 Substitutions= 10: (C 35S) (C 65S) (C 89S) (T137F)  
 (C195S) (L202G) (M259L) (L261N) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCGCGGACGGGCACAAAGGTTTAGCCACGATTCCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTctGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAACCCAAATGtctGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCAATGCTTGGCAAAGACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACctcGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTACGAGTCCTTATCGTctcTCCGAGATCACTGCTATqggaTTCCATGGACCCCAATGAGAACCACCTTGACTCAC  
 TCGTCGCTCAAGCTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCCTCCTTGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGGCActgAagaaCCTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAAAGAACATAGAGAACGCGAGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCAACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAACCTGATTAGCGCTgctGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGATACAACCGGTGAAGGTTTGGATtcdGGTG  
 TTTTGTGTTGGATTGGACCGGTTATGACTGTGTGAGACTGTGTTCTTCGTAGCGTCCCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 15 Substitutions= 10: (C 35S) (C 65S) (C 89S) (C195S)  
 (I201F) (L202G) (M259L) (L261N) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCGCGGACGGGCACAAAGGTTTAGCCACGATTCCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTctGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC

FIG. 12 (continued)

TATCTGCAAGAGAAACCCAAACAATG[ctct]GAGTTTCATGGCTCCATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGGT  
 CCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCIGCACCAACGCTGGCTTGACATGCCCCGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTCAGAGTCCTTATCGT[ctcc]TCCGAGATCACTGCT[ttcggg]TTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTGCCGACACTGAGAAGGCA[ctg]AAG[aaac]CACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGATCCCTTGATGTCGCAAGAACATAGAGAACGCGAGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCAACCCAGGTGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCT[gtct]GT  
 GTTGTTCATCATTCACGAGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGAT[ctcc]GGTG  
 TTTTGTGTTGGATTGGACCGGGTAAGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 16 Substitutions= 10: (C 35S) (C 65S) (C 89S) (T137F)  
 (C195S) (L202G) (M259L) (L261N) (C346A) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCGGACGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[ctcc]GTCCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCA[ttct]GAGAAACACGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATG[ctct]GAGTTTCATGGCTCCATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGGT  
 CCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCAC[ctcc]GCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTCAGAGTCCTTATCGT[ctcc]TCCGAGATCACTGCTAT[ctggg]TTCCATGGACCCCAATGAGAACCACTTGACTCAC

## FIG. 12 (continued)

TCGTCCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGGCCGTAGAAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCCGACACTGAGAAAGCA[ctg]AAG[aaac]CACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTAACCTTGATGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTTGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAAGGTGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCT[gct]GT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGTGAAGTTTGGAT[tc]cGGTG  
 TTTTGTGTGGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 17 Substitutions= 10: (C 35S) (C 65S) (C 89S) (C195S) (L202G) 63/150  
 (M259L) (L261N) (L268M) (C346A) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCTGTGAGCCGGACGGGCACAAGGTTTAGCCACGATTCTTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[tc]cGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCAT[tc]cTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAATG[tc]cTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGGT  
 CCCAATGCTTGGCAAAGAACCCGACGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAAGCACCTTGTCTGAAAACAACAA  
 GGGCTACGAGTCTTATCGT[tc]cTCCGAGATCACTGCTAT[ggg]TTCCATGGACCCCAATGAGAACCACCTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCA[ctg]AAG[aaac]CACTTGAGAGAGGGAGG  
 Gatg[acg]TTTCAGTTGCATAGAGATGTACCTTGTATGGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAAGGTGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCT[gct]GT

## FIG. 12 (continued)

GTGTTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGATACAACCGGTGAAGGTTTGGATtccGGTG  
 TTTTGGATTTGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 18 Substitutions= 3: (C 35S) (C 89S) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGATCTTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTAAACGCAATTGTGAGAAACACAGCGATAAGAAACGATACCTAGCCCTACCCGAAGAC  
 TATCTGCAAGAAACCCAAATGtctGAGTTTCATGGCTCCATCCTTAAACGCTCGACACAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAATCAGATCACCACCTCATCT  
 TCTGCACCACGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGGCGACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTACAGAGTCCTTATCGTCTGTCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCAACCCAGGTGGTCGAGCCATATTGGACCAAGGTGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGATACAACCGGTGAAGGTTTGGATtccGGTG  
 TTTTGGTTTGGATTTGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 19 Substitutions= 4: (C 35S) (C 89S) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGCGCGGACGGGCACAAAGGTTTAGCCACGATCTTTGCCAT

65/150

## FIG. 12 (continued)

TGGCACTGCTACTCCTCCCAAT[cc]GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGt[ct]GAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGGCGACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCAAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACCGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATGGAACCTCAGTTTCTGGATGGTGACCCAGTGCTGAGCCCATATTGACCAAGGTGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAACCTGATTAGCGTg[ct]GT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGATACAACCGGTGAAGTTTGGATt[cc]GGTG  
 TTTTGTGTTGGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 20 Substitutions= 4: (C 35S) (C 65S) (C 89S) (C 372S)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[cc]GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTT[ct]GAGAAACACGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGt[ct]GAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGGCGACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA

66/150

## FIG. 12 (continued)

GGGCTACGAGTCCTTATCGCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTCAGGCCCTCACTTGCCGCTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGGTCGAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTCACCCAGGTGGTCGAGCCATATTGACCAGGTGGAGCCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTTGT  
 GTTGTTCATCAATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGAATTCGGTG  
 TTTTGTGTTGGATTTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 21 Substitutions= 5: (C 35S) (C 89S) (C195S) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGCAACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAATGCTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCAICT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCGCTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGGTCGCAAGAAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCAACCCAGGTGGTCGAGCCATATTGACCAGGTGGAGCCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTGCTGT

## FIG. 12 (continued)

GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATtccGGGTG  
 TTTTGTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAGCTGA

> 2-PS variant= 22 Substitutions= 6: (C 35S) (C 65S) (C 89S) (C195S) (C346A)  
 (C372S)

ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTCTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCAATtctGAGAAACACAGCGATAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAATGtctGAGTTTCATGGCTCCATCCTTAAACGCTCGACAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAATCCAAATCACCCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCCCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCAGAGTCCTTATCGTctccTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGTGCCAAAGAACATAGAGAACCGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGACCAAGGTGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTgctGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATtccGGGTG  
 TTTTGTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAGCTGA

67/150

68/150

## FIG. 12 (continued)

> 2-PS variant= 23 Substitutions= 2: (C 35S) (T137F)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCGCGGACGACAAAGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCAC**ct**GGTGGCGTTGACATGCCCGGTGCTGACTATCAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCCTCCGGTAGCCAAAGGACCTTGCTGAAACAAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGCACTGAGAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGTCGCCAAAGAACATAGAGAACCGCAGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCCACCCAGGTGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 24 Substitutions= 2: (C 35S) (T137M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCGCGGACGACAAAGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

69/150

FIG. 12 (continued)

TCATGACCGatgGCTGGCGTTGACATGCCGGTGCTGACTATCAACTCGTCAAACTCCTTTGGTCTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCGCCGGGCACAGTCCTCCGGCTAGCCAAAGACCTTGTCTGAAACAACAA  
 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGAACCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGGTCGCAAGAACATAGAGAACGCAGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCCACCCAGGTGGTCGAGCCATATTGGACCAAGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTCACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTGGATTGGACCGGTAIGACTGTGTGAGACTGTGTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 25 Substitutions= 2: (C 35S) (M259F)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATccgGTGCGTCAAGCTGATATGCAGACTATTATTTCGTGTCACATAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCGGTGTGACTATCAACTCGTCAAACTCCTTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCGCCGGCGCACAGTCTCCGGCTAGCCAAGGACCTTGTCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCCGACACTGAGAAGGCAtttAAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGGTCGCAAGAACATAGAGAACGCAGCGAGAAAGCGTTGTCTC

## FIG. 12 (continued)

CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTCACGAGGTGAGGAAGAGATCTATGGCGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 26 Substitutions= 2: (C 35S) (T137L)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTTCGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCACAACAAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGGT  
 CCCAATGCTTGGCAAAGAAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACcctdGCTGGCGTTGACATGCCCCGGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGAACCAATGAGAACCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTGTCAACTGATCAAAACAATCTTGGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACAATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGAATGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTCACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

71/150

**FIG. 12 (continued)**

> 2-PS variant= 27 Substitutions= 2: (C 35S) (I343F)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCGCGGACAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**TCG**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCACAATGTGTGAGTTCATGGCTCCATCCTTAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGCAAGAGCCGAGTCAAGGCCATIGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCGCCGGCCAGTCCTCCGGCTAGCCCAAGGACCTTGCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGAACCAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTATTCGGAGACGGAGCTGCAGCACTCATTTGIGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGGTCGCAAGAACATAGAGAACGCGAGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTTCGTGATGGTGCAACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAACATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTG**ttt**AGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 28 Substitutions= 2: (C 35S) (I343M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCGCGGACAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**TCG**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCACAATGTGTGAGTTCATGGCTCCATCCTTAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGCAAGAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

## FIG. 12 (continued)

TCTGCACACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTAATGTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCCTTATCGTCTGCTCCGAGATCAGTCTATCTTATTCATGGAACCAATGAGAAACCACTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAAACGGCCA  
 ATATTCCGAGATCGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGACCCCTTGATGGTCGCAAGAACAATAGAGAACGCGAGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAAGTGGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGatgAGCGCTTGTGT  
 GTTGTTCATCATTCGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGAGTACAACCCGGTGAAGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 29 Substitutions= 2: (C 35S) (I201F)

ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCGTGAGGCCCGGACCGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTCCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACCGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGT  
 CCCAATGCTTGGCAAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCGACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCCTTATCGTCTCGAGATCACTGCTttdTTATTCCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCCATAGAGATGTACCCCTTGATGGTCTGCCAAAGAACATAGAGAACGCGAGGAGAAAGCGTTGTCTC

## FIG. 12 (continued)

CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCCGAAGGAAGAGTACAACCGGTGAAGGTTTGATTGCGGTG  
 TTTTGTGTTGGATTTGACCCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 30 Substitutions= 2: (C 35S) (C169A)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTTAAACGCATTTGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGAgctGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCAATAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCCGAAGGAAGAGTACAACCGGTGAAGGTTTGATTGCGGTG  
 TTTTGTGTTGGATTTGACCCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

74/150

## FIG. 12 (continued)

> 2-PS variant= 31 Substitutions= 2: (C 35S) (L268M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGCACAAAGGTTAGCCACGATCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCGCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAACAACAA  
 GGGCTCACGAGTCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACCGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCCTCCTGAGAACCGGCGCA  
 ATATTCGAGATCGTGTCAACTGATCAACAATCTTCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
**gat**ACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCGAGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTCGATGGTCCACCCAGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGCGAAGGGAAGAGTACAAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 32 Substitutions= 2: (C 35S) (L202F)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGCACAAAGGTTAGCCACGATCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

75/150

## FIG. 12 (continued)

TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGAGATCACTGCTATC**ttc**ttccatggaaccatgagaaccaccttgactcac  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGGTCGCAAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGATTCGGGTG  
 TTTTGTGTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 33 Substitutions= 2: (C 35S) (I201M)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCCGGACGGGCAACAAGGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAAT**ccc**gtcgcctcaagctgattatgcagactattattttcgtgtcactaagagcggaacata  
 TGGTTGATCTTAAGAGAAATTTAAACGCATTTGTGAGAAAACAGCGATAAAGAAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGAGATCACTGCT**atg**ttattccatggaccccaatgagaaccaccttgactcac  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCTTGATGGTCGCAAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC

76/150

## FIG. 12 (continued)

CACTAGCGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGTGGTCGAGCCATATTTGGACCAGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATTCATTGACGAGGTGAGGAAGAGATCTATGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCGTCGCGTTACTGCTGCGGTGCGCAAT  
 GGAAACTGA

> 2-PS variant= 34 Substitutions= 2: (C 35S) (I201L)

ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCGTGAGGCCGGACGAGCAAGGTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[cc]GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCAAGTCCCTCCGGCTAGCCAAAGGACCTTGCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCT[ctg]TTATTCCATGGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCTGCAAGAACATAGAGAACGCGAGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCAACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATTCATTGACGAGGTGAGGAAGAGATCTATGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCGTCGCGTTACTGCTGCGGTGCGCAAT  
 GGAAACTGA

77/150

## FIG. 12 (continued)

> 2-PS variant= 35 Substitutions= 2: (C 35S) (L202M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCTGTAGGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTACCCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTCAACGGCGT  
 CCCAATGCTTGGCAAAGAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATcctgTTCCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAACAATCTTCCCGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACGACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAAGGTGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 36 Substitutions= 4: (C 35S) (L202G) (M259L) (L261N)  
 ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCTGTAGGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTACCCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTCAACGGCGT  
 CCCAATGCTTGGCAAAGAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

## FIG. 12 (continued)

TCTGCACCACGCTGGCGTTGACATGCCCGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAGGACCTTGTGTAACAAACAA  
 GGGCTACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTCAAGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCActgAAGaadCACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGTCGCAAGAACATAGAGAACGCGAGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTTGATGGTGACCCAGGTGGTCGAGCCATATTGACCAAGGTGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

78/150

> 2-PS variant= 37 Substitutions= 9: (C 35S) (C 65S) (C 89S) (C195S) (L202G)

(M259L) (L261N) (C346A) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATgtctGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTCAACGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCACCGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTtccTCCGAGATCACTGCTATtgggaTTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCActgAAGaadCACTTGAGAGAGGGAGG

79/150

## FIG. 12 (continued)

GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGACCAAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTgctGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGTTTGGATtccGGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 38 Substitutions= 2: (C 35S) (L202G)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCAACAATGTGTGAGTTCATGGCTCCATCCTTAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAGAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATcggattCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGTTTGGATTCCGGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

## FIG. 12 (continued)

> 2-PS variant= 39 Substitutions= 3: (C 35S) (M259L) (L261N)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAGGACCTTGTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCCTACTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTGTCAACTGATCAAAACAATCTTGGCGGACACTGAGAAAGCActgAAgaadCACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACAATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCAACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCCGGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 40 Substitutions= 7: (C 35S) (C 65S) (C 89S) (C195S)  
 (L202G) (C346A) (C372S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATtctGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTctGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT

FIG. 12 (continued)

CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGGCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTCCGAGATCACTGCTATGGATTCCATGAGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCCTTATTTCGGAGACGGAGCTGCAGCACTCAATTGTGGTTCAAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACAATAGAGAACGCCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTgctGT  
 GTTGTTCATCAATTGACGAGGTGAGGAAGAGATCTATGGCGAAGGAAGAGTACAACCCGGTGAAGGTTTGGATCCdGGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 41 Substitutions= 1: (C 35S)

ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCTGTAGGCGCGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCAAHtccGTGCTCAAGCTGATTATGCAGACTATTATTTCGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAACCAACAATGTGTGAGTTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGCAAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTGTCAACTGATCAAAACAATCTTGCCCGCACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG

## FIG. 12 (continued)

GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAAAGAACATAGAGAACGCAGCGGAGAAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTTGACCAAGGTGGAGCGGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGTACAACCGGTGAAGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 42 Substitutions= 4: (C 35S) (T137L) (L202F) (L268M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATTCTGGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTTCTGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTACCGGCGGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCTCTGGCTGGCTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGCCTGAAAACAACAA  
 GGGCTCACGAGTCTTATCGTCTGCTCCGAGATCAGTCTATCTCTTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GATGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCAGCGGAGAAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTTGGACCAAGGTGGAGCGGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

## FIG. 12 (continued)

> 2-PS variant= 43 Substitutions= 5: (C 35S) (T137M) (L202F) (L268M) (I343F)  
 ATGGGATCTTACTATCCGATGATGTGGAGGTGATTCGTGAGCGCGGACGACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCGTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAAACCAACAATGTGTGAGTTTCAATGGCTCCATCCTTAAACGCTCGACAAAGACCTAGTGTCACCGGCGGT  
 CCCAATGCTTGGCAAGAACGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCAC**atd**GCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCGGCGGACAGTCTCCGCTAGCCAAAGGACCTTGCTGAAACAAACAA  
 GGGCTCACGAGTCTTATCGTCTGCTCCGAGATCACTGCTAT**ctc**TTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
**gat**ACGTTTCAGTTGCATAGAGATGTACCCCTGATGGTCGCAAGAACATAGAGAACGACGCGGAGAAAGCGTTGTCTC  
 CACTAGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCAACCCAGGTGGTCAGCCATATTGGACCAAGGTGGAGCGA  
 AAACCTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTG**ttt**AGCGCTTGTT  
 GTTGTTCATCATTCACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGATACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCCGGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 44 Substitutions= 4: (C 35S) (L202F) (L268M) (I343F)  
 ATGGGATCTTACTATCCGATGATGTGGAGGTGATTCGTGAGCGCGGACGCGACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCGTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCCGAAGAC  
 TATCTGCAAGAGAAACCAACAATGTGTGAGTTTCAATGGCTCCATCCTTAAACGCTCGACAAAGACCTAGTGGTCAACCGCGT

## FIG. 12 (continued)

CCCAATGCTTGGCAAAGAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGACCAACCGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAACAACAA  
 GGGCTACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATC**ttc**ttccatggaccccaatgagaaaccacttgactcac  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGGGAGG  
 gatgacgTTTCAGTTGCATAGAGATGTACCTTGATGGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAAGGTGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTG**ttt**agcgcTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGATACAACCGGTGAAGTTTGGATTGCCGTG  
 TTTTGTGTTGGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTTCGTAGCGTCCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 45 Substitutions= 2: (C 35S) (C372S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCGCGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcg**gtcgcTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGGT  
 CCCAATGCTTGGCAAAGAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAACAACAA  
 GGGCTACGAGTCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGGGAGG

85/150

## FIG. 12 (continued)

GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGACCAAGGTGAGCGGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGTTTGGAT<sup>tc</sup>GGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 46 Substitutions= 2: (C 35S) (C372A)

ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGGCCGGACGGGCACAAAGTTTAGCCACGATTCTTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT<sup>tc</sup>GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTTAAACGCATTTGTGAGAAACACGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGGT  
 CCCAATGCTTGGCAAAGAACCCGCAGTCAAGGCCATTGATGAATGGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACCAACAA  
 GGGCTACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCGTAGAACGGCCA  
 ATATTTCGAGATCGTGTCAACTGATCAACAATCTTGGCCGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGTGATGGTCCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGACCAAGGTGAGCGGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGTTTGGAT<sup>gc</sup>GGTG  
 TTTTGTGTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

## FIG. 12 (continued)

> 2-PS variant= 47 Substitutions= 2: (C 35S) (C346S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCTGTGAGCCGGACGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[**tc**]GTGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAATCACCCACCTCATCT  
 TCTGCAACCACCGCTGGCGTTGACATGCCCCGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCAGAGTCTTATCGTCTCGAGATCACTGCTATCTTATTTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCCGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTAACCTTGATGGTCGCAAGAACATAGAGAACGACGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCT[**tc**]GT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGAACCGGATGACTGTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 48 Substitutions= 2: (C 35S) (C346A)  
 ATGGGATCTTACTCATCCGATGATGTGGAGTGATTCTGTGAGCCGGACGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[**tc**]GTGCTCAAGCTGATTATGACAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGGAATAACCAAAATCCAAATCACCCACCTCATCT

87/150

## FIG. 12 (continued)

TCTGACCAACCGCTGGCGTTGACATGCCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGAGATGTGCCCGCGGCACAGTCTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTACGAGTCCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGGCCGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAAACATAGAGAACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTCGATGGTGCACCCAGGTGGTCGAGCCATAATTGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCCGTgctGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCCGGTG  
 TTTTGTGTTGGATTTGGACCGGGIATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 49 Substitutions= 2: (C 35S) (C195S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGCCCGACGGCACAAAGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATtcdGTCTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCATAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGAGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTACGAGTCCCTTATCGTctccGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGGCCGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCTGCAAGAACATAGAGAACGCGGAGAAAGCGTTGTCTC

## FIG. 12 (continued)

CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCAGCCATATTGGACCAGGTGAGCGGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCCGAAGGAAGATACAACCGTGAAGGTTTGGATTGCCGGTG  
 TTTTGTGTTGGATTGGACCCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCCGGTTACTGCTGCCGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 50 Substitutions= 2: (C 35S) (C195A)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT[ccc]GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAAACCCAAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACACAGACCTAGTGTACCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCCGCTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCCGGCGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGTGTAACAAACAA  
 GGGCTCACGAGTCCCTTATCGTC[gcc]TCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACCGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCCGACACTGAGAAAGCAATGAATTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTGCGAAAGAACATAGAGAACCGCAGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCAGCCATATTGGACCAGGTGGAGCGGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCCGAAGGAAGATACAACCGTGAAGGTTTGGATTGCCGGTG  
 TTTTGTGTTGGATTGGACCCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCCGGTTACTGCTGCCGTTGCCAAT  
 GGAAACTGA

## FIG. 12 (continued)

> 2-PS variant= 51 Substitutions= 2: (C 35S) (C135S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**CGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT  
 TC**tc**ACCACCGCTGGCTTGACATGCCCGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAACAACAA  
 GGGCTACGAGTCCCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGGCGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCTTGATGTGCGAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCAGCCATATTGGACCAGGTGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 52 Substitutions= 2: (C 35S) (C135A)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**CGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAAGATCACCCACCTCATCT

FIG. 12 (continued)

TCgcdACCACCGCTGGCGTTGACATGCCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTCTGTGAAAACAACAA  
 GGGCTCAGAGTCCCTTATCTGCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGGTTTCAGGCCCTCACTTGCCCGTAGAACCGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACATAGAGAACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAAAGTGAATGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTTGGATTGGACCGGTATGACTGTTGAGACTGTGTTCTTCGTAGCGTCCGGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 53 Substitutions= 2: (C 35S) (C 89S)  
 ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCGTGAGGCCGGACGGGCACAAGGTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATccdgTCGCTCAAGCTGATTATGCAGACTATTATTTCTGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTAAACCGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAACAATGtctGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAAGAAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATAAGTTGTACCAACAGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAGGACCTTGTGTAACAAACAACAA  
 GGGCTCACGAGTCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG

## FIG. 12 (continued)

GTTGACGTTTCAGTTGCAATAGAGATGTACCCCTTGATGTCGCAAGAACATAGAGAACGCGAGAGAAAGCGTTGTCTC  
 CACTAGGGATAAAGTGAATGGAACTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTGGATTGGACCGGGTATGACTGTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 54 Substitutions= 2: (C 35S) (C 89A)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAGGTTAGCCACGATCTTGCCAT  
 TGGCACTGCTACTCCTCCCAATccGGTCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACATAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAATTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGgctGAGTTTCATGGCTCCATCCTTAACGCTCGACAAGACCTAGTGGTCACCGGGGT  
 CCCAATGCTTGGCAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATAATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAACAACAA  
 GGGCTCAGAGTCCTTATCGTCTCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTCAGTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTCCCGACACTCAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGTCGCAAGAACAATAGAGAACGCGAGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

92/150

## FIG. 12 (continued)

> 2-PS variant= 55 Substitutions= 2: (C 35S) (C 65S)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGTTTAGCCACGATCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcct**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCAT**tcct**GAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATAGTTGTACCAACAGGGATGTGCCCGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAAACAA  
 GGGCTCACGAGTCCTTATCGTCTCGTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCCGACACCGAGCTGCAGCACTCATTTGTGGTTCAAGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGACCCCTTGATGGTCGAAAGAACATAGAGAACGACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGCCGGAAGGAAGAGTACAACCCGGTGAAGGTTTGGATTGCCGGTG  
 TTTTGTGTGGATTGGACCGGTAATGACTGTTGAGACTGTTGTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 56 Substitutions= 2: (C 35S) (C 65A)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGCACAAAGTTTAGCCACGATCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcct**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCAT**gct**GAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGT  
 CCCAATGCTTGGCAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCACT

## FIG. 12 (continued)

TCTGCACCACCGCTGGCGTTGACATGCCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGATGTGCCCGCGGACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTACGAGTCCTTATCGCTCGTCCGAGATCACTGCTATCTTATCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAAACATAGAGAACGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACCTCAGTTTCTGGATGGTGACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCAATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAAGAGTACAACCGGTGAAGTTTGGATTGCGGTG  
 TTTTGGTTTGGATTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGCGGTTGCCAAT  
 GGAACACTGA

> 2-PS variant= 57 Substitutions= 2: (C 35S) (C169S)

ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCGGACGGCAACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT~~tc~~GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACCGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCACCGCTGGCGTTGACATGCCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGA~~tc~~GTGCCCGCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCAATTGTGGTTTCAGGCCCTCACTTGCCCGTAGAACGGCCA  
 ATATTCCGAGATCGTCAACTGATCAAAACAATCTTCCCGACACTGAGAAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAAACATAGAGAACGCGGAGAAAGCGTTGTCTC

## FIG. 12 (continued)

CACTAGGGATAAAGTATTGGAACCTCAGTTTCTGGATGGTGCACCCAGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTCTTAGTGAATACGGAACCTGATTAGCGCTTGTGT  
 GTTGTTCAATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGATCAACCCGTTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTGGACCCGGTATGACTGTTGAGACTGTTGTTCTTTCGTAGCGTCCGCTTACTGCTGCCGTTGCCAAT  
 GGAAGACTGA

> 2-PS variant= 58 Substitutions= 2: (C 35S) (N341H)

ATGGGATCTTACTCATCCGATGATGTGAGGTGATTCTGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATCTTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tc**GTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAGAGAAAATTAAACGCATTTGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCAAATGTGTGAGTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAAGAAGCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCCGGTGTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCCGGCGCACAGTCCCTCCGGCTAGCCAAAGGACCTTGTGAAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCATGGACCCAAATGAGAACCCACCTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTGCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTGCCAAAGAAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTC  
 CACTAGGGATAAAGTATTGGAACCTCAGTTTCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AAATAAACCTTAAGGAAGATAAGTTAAGGCTAGCAGGCATGTGCTTAGTGAATACGG**cad**CTGATTAGCGCTTGTGT  
 GTTGTTCATCATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGATCAACCCGTTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTGGACCCGGTATGACTGTTGAGACTGTTGTTCTTTCGTAGCGTCCGCTTACTGCTGCCGTTGCCAAT  
 GGAAGACTGA

95/150

## FIG. 12 (continued)

> 2-PS variant= 59 Substitutions= 2: (C 35S) (H308Q)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGACGGGCACAAGGTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcg**GTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAGACCTTGTCTGAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACGTCTATCTTATTCATGGACCCCAATGAGAACCCCTTGACTCAC  
 TCGTCCGCTCAAGCTTTATTCGGGACACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCGAGATCGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GTTGACGTTTCAGTTGCATAGAGATGACCCCTTGATGGTCGAAAGACATAGAGAACGACGCGAGAAAGCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTCGATGGT**cag**CCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCGCGAAGGGAAGAGTACAACCCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTGTGGATTGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 60 Substitutions= 3: (C 35S) (L202F) (L268F)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGACGGGCACAAGGTTTAGCCACGATTCTTGCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcg**GTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAACCCCAACAATGTGTGAGTTTCATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCCGGCGT  
 CCCAATGCTTGGCAAAGAACCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT

## FIG. 12 (continued)

TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAA  
 CGCTATAIGTTGTACCAACAGGATGTGCCGCCGGCACAGTCCCTCCGGCTAGCCAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCCTTATCGTCTGCTCCGAGATCACCTGCTATC**ttc**TTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 G**ttc**ACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACAATAGAGAACGCGAGAAAGCGTTGTCTC  
 CACTAGGATAAAGTGAATGGAACTCAGTTTCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAAGGTGGAGCGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGCCGGAAGGAAGAGTACAACCCGTTGAAGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 61 Substitutions= 3: (C 35S) (L268M) (I343M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTGAGGCCCGGACGGGCACAAAGGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAAT**tcg**GTGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACCTAAGAGCGAACATA  
 TGGTTGATCTTAAAGAGAAATTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCAAACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAAGACCTAGTGGTCAACCGCGT  
 CCCAATGCTTGGCAAAGAAGCCGCAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAA  
 CGCTATAIGTTGTACCAACAGGATGTGCCGCCGGGCACAGTCCCTCCGGCTAGCCAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCCTTATCGTCTGCTCCGAGATCACCTGCTATCTTATTCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTCCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGCAATGAAGTTACACTTGAGAGAGGGAGG  
 G**atg**ACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCAAGAACAATAGAGAACGCGAGAAAGCGTTGTCTC

97/150

## FIG. 12 (continued)

CACTAGGGATAACTGATTGGAACTCAGTTTCTGGATGGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AACTAAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGatgAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

> 2-PS variant= 62 Substitutions= 3: (C 35S) (L202F) (L268M)  
 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCCAT  
 TGGCACTGCTACTCCTCCCAATtccGTGCGCTCAAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGGAACATA  
 TGGTTGATCTTAAAGAGAAATTTAAACGCATTTGTGAGAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGAC  
 TATCTGCAAGAGAACCCACAATGTGTGAGTTCAATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGTGTACCGGCGT  
 CCCAATGCTTGGCAAAGAACCCGAGTCAAGGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCT  
 TCTGCACCAACCGCTGGCGTTGACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCTTGGTCTCTCCCCCTTCAGTCAAA  
 CGCTATATGTTGTACCAACAGGGATGTGCCCGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTCTGAAAACAACAA  
 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtccTTCCATGGACCCCAATGAGAACCACTTGACTCAC  
 TCGTCGCTCAAGCTTTATTCCGGAGACGGAGCTGCAGCACTATTGTGGGTTTCAGGCCCTCACTTGGCCGTAGAACGGCCA  
 ATATTTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGG  
 GatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTCGCCAAAGAACATAGACAACGCGGAGAAACCGTTGTCTC  
 CACTAGGGATAACTGATTGGAACTCAGTTTCTGTGATGTGCACCCAGGTGTCGAGCCATATTGGACCAGGTGGAGCGGA  
 AAACATAACCTTAAGGAAGATAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGT  
 GTTGTTCATCATTTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGATACAACCGGTGAAGGTTTGGATTGCGGTG  
 TTTTGTTTGGATTTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAAT  
 GGAAACTGA

FIG. 13

Nucleotide sequence alignments: 2-ps

|       |    |                                                                        |    |
|-------|----|------------------------------------------------------------------------|----|
| > Seq | 1  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA | 70 |
| > Seq | 2  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 3  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 4  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 5  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 6  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 7  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 8  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 9  | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 10 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 11 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 12 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 13 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 14 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 15 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 16 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 17 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 18 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 19 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 20 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |
| > Seq | 21 | ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCTGTAGGCCGGACGGGCACAAGGTTTAGCCACGA |    |

## FIG. 13 (continued)

> Seq 22 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 23 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 24 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 25 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 26 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 27 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 28 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 29 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 30 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 31 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 32 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 33 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 34 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 35 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 36 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 37 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 38 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 39 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 40 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 41 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 42 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 43 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 44 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA  
> Seq 45 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTTAGCCACGA

FIG. 13 (continued)

> Seq 46 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 47 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 48 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 49 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 50 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 51 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 52 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 53 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 54 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 55 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 56 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 57 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 58 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 59 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 60 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 61 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA  
 > Seq 62 ATGGGATCTTACTCATCCGATGATGTGGAGGTGATTCGTGAGCCGGACGGGCACAAAGGTTAGCCACGA

71 . . . . . 140  
 > Seq 1 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 2 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 3 TTCTTGCCATTGGCACTGCTACTCCTCCCAATgccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 4 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 5 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG

## FIG. 13 (continued)

> Seq 6 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 7 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 8 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 9 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 10 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 11 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 12 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 13 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 14 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 15 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 16 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 17 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 18 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 19 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 20 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 21 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 22 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 23 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 24 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 25 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 26 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 27 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 28 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 29 TTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG

## FIG. 13 (continued)

> Seq 30 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 31 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 32 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 33 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 34 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 35 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 36 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 37 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 38 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 39 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 40 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 41 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 42 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 43 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 44 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 45 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 46 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 47 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 48 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 49 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 50 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 51 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 52 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
> Seq 53 TTCTTGCCATTGGCACTGCTACTCCTCCCAATtccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG

103/150

## FIG. 13 (continued)

> Seq 54 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 55 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 56 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 57 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 58 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 59 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 60 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 61 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG  
 > Seq 62 TTCTTGCCATTGGCACTGCTACTCCTCCCAATccGTCGCTCAAGCTGATTATGCAGACTATTATTTTCG

141

.

.

.

.

.

.

210

> Seq 1 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 2 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 3 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 4 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 5 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 6 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 7 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 8 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 9 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 10 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTctGAGAAAACAGCGATA  
 > Seq 11 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTGTGAGAAAACAGCGATA  
 > Seq 12 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTctGAGAAAACAGCGATA  
 > Seq 13 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCCATTTctGAGAAAACAGCGATA

## FIG. 13 (continued)

> Seq 14 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 15 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 16 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 17 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 18 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 19 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 20 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 21 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 22 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 23 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 24 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 25 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 26 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 27 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 28 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 29 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 30 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 31 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 32 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 33 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 34 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 35 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 36 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA  
> Seq 37 TGTCACCTAAGAGCGGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATTtctGAGAAAAACAGCGGATA

## FIG. 13 (continued)

> Seq 38 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 39 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 40 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTCTGAGAAAACAGCGGATA  
> Seq 41 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 42 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 43 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 44 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 45 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 46 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 47 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 48 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 49 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 50 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 51 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 52 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 53 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 54 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 55 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTCTGAGAAAACAGCGGATA  
> Seq 56 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 57 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 58 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 59 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 60 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA  
> Seq 61 TGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAAATTTAAACGCATTTGTGAGAAAACAGCGGATA

FIG. 13 (continued)

|       |     |        |        |        |        |        |        |        |        |        |        |        |           |
|-------|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----------|
| > Seq | 62  | TGTCAC | TAAGAG | CGAACA | TATGGT | TGATCT | TAAAGA | GAATTT | AAACGC | ATTTGT | GAGAAA | CACG   | GATA      |
|       |     |        |        |        |        |        |        |        |        |        |        |        |           |
|       | 211 | .      | .      | .      | .      | .      | .      | .      | .      | .      | .      | 280    |           |
| > Seq | 1   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 2   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 3   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 4   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 5   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 6   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 7   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | GTGAGT | TCATGGCTC |
| > Seq | 8   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 9   | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 10  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 11  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 12  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 13  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 14  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 15  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 16  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 17  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 18  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 19  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 20  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |
| > Seq | 21  | AAGAA  | ACGATA | ACCTAG | CCCTCA | CCGAAG | ACTATC | TGCAAG | AGAA   | CCCCAA | CAATGT | ctGAGT | TCATGGCTC |

## FIG. 13 (continued)

> Seq 22 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 23 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 24 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 25 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 26 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 27 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 28 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 29 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 30 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 31 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 32 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 33 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 34 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 35 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 36 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 37 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 38 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 39 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 40 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 41 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 42 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 43 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 44 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC  
> Seq 45 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAAACCCAAATGctGAGTTCATGGCTC

108/150

## FIG. 13 (continued)

> Seq 46 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 47 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 48 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 49 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 50 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 51 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 52 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 53 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGtctGAGTTCATGGCTC  
 > Seq 54 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGgctGAGTTCATGGCTC  
 > Seq 55 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 56 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 57 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 58 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 59 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 60 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 61 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC  
 > Seq 62 AAGAAACGATACCTAGCCCTCACC GAAGACTATCTGCAAGAGAACCCCAACAATGTGTGAGTTCATGGCTC

281 . . . . . 350  
 > Seq 1 CATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGTCCCAATGCTTGGCAAGAAGCCGCAGTCAA  
 > Seq 2 CATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGTCCCAATGCTTGGCAAGAAGCCGCAGTCAA  
 > Seq 3 CATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGTCCCAATGCTTGGCAAGAAGCCGCAGTCAA  
 > Seq 4 CATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGTCCCAATGCTTGGCAAGAAGCCGCAGTCAA  
 > Seq 5 CATCCTTAAACGCTCGACAAGACCTAGTGGTCAACCGCGTCCCAATGCTTGGCAAGAAGCCGCAGTCAA

## FIG. 13 (continued)

> Seq 6 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 7 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 8 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 9 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 10 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 11 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 12 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 13 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 14 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 15 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 16 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 17 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 18 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 19 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 20 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 21 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 22 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 23 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 24 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 25 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 26 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 27 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 28 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA  
> Seq 29 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGCGTCCCAATGCTTTGGCAAAGAGCCGCAGTCAA

## FIG. 13 (continued)

> Seq 30 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 31 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 32 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 33 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 34 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 35 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 36 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 37 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 38 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 39 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 40 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 41 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 42 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 43 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 44 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 45 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 46 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 47 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 48 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 49 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 50 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 51 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 52 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA  
> Seq 53 CATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGGTCCCAATGCTTGGCAAAGAGCCGCAGTCAA

## FIG. 13 (continued)

> Seq 54 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 55 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 56 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 57 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 58 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 59 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 60 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 61 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA  
 > Seq 62 CATCCTTAAACGCTCGACAAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAAGAGCCGCAGTCAA

111/150

351

> Seq 1 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 2 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 3 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 4 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 5 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 6 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 7 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 8 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 9 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 10 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 11 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 12 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT  
 > Seq 13 GGCCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCAACCGCTGGCGTT

420

## FIG. 13 (continued)

> Seq 14 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCTtcGCTGGCGTT  
> Seq 15 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 16 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCTtcGCTGGCGTT  
> Seq 17 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 18 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 19 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 20 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 21 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 22 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 23 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCTtcGCTGGCGTT  
> Seq 24 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCatgGCTGGCGTT  
> Seq 25 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 26 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCTtcGCTGGCGTT  
> Seq 27 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 28 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 29 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 30 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 31 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 32 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 33 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 34 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 35 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 36 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 37 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT

## FIG. 13 (continued)

> Seq 38 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 39 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 40 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 41 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 42 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 43 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 44 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 45 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 46 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 47 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 48 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 49 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 50 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 51 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 52 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 53 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 54 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 55 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 56 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 57 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 58 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 59 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 60 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT  
> Seq 61 GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACACCGCTGGCGTT

|        |     |                                                                        |     |
|--------|-----|------------------------------------------------------------------------|-----|
| >> Seq | 62  | GGCCATTGATGAATGGGACTACCAAAATCCAAGATCACCCACCTCAICTTCTGCACCAACCGCTGGCGTT |     |
| >> Seq | 421 | . . . . .                                                              | 490 |
| >> Seq | 1   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 2   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 3   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 4   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 5   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 6   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 7   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 8   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 9   | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 10  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 11  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 12  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 13  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 14  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 15  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 16  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 17  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 18  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 19  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 20  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |
| >> Seq | 21  | GACATGCCCGGTGCTGACTATCAACTCGTCAAACCTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT |     |

## FIG. 13 (continued)

> Seq 22 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 23 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 24 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 25 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 26 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 27 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 28 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 29 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 30 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 31 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 32 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 33 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 34 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 35 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 36 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 37 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 38 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 39 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 40 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 41 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 42 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 43 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 44 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT  
> Seq 45 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCTTCAGTCAAACGCTATATGT

116/150

## FIG. 13 (continued)

```

> Seq 46 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 47 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 48 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 49 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 50 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 51 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 52 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 53 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 54 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 55 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 56 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 57 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 58 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 59 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 60 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 61 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT
> Seq 62 GACATGCCCGGTGCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCCTTCAGTCAAACGGCTATATGT

491 . . . . . 560
> Seq 1 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA
> Seq 2 TGTACCAACAGGGAGctGCCGCCGGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA
> Seq 3 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA
> Seq 4 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA
> Seq 5 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAACAACAA

```

117/150

## FIG. 13 (continued)

> Seq 6 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 7 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 8 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 9 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 10 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 11 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 12 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 13 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 14 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 15 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 16 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 17 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 18 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 19 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 20 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 21 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 22 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 23 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 24 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 25 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 26 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 27 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 28 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA  
> Seq 29 TGTACCAACAGGGATGTGCCGCCGGCGGCACAGTCTCCGGCTAGCCAAAGGACCTTGTGAAACAACAA

## FIG. 13 (continued)

> Seq 30 TGTACCAACAGGAGctGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 31 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 32 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 33 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 34 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 35 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 36 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 37 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 38 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 39 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 40 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 41 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 42 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 43 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 44 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 45 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 46 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 47 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 48 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 49 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 50 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 51 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 52 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 53 TGTACCAACAGGGATGTGCCGCCGGCGCACAGTCCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA

FIG. 13 (continued)

> Seq 54 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 55 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 56 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 57 TGTACCAACAGGGATctGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 58 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 59 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 60 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 61 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA  
> Seq 62 TGTACCAACAGGGATGTGCCGCCGGGCGCACAGTCTCCGGCTAGCCAAAGGACCTTGCTGAAAAACAACAA

561 . . . 630  
> Seq 1 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 2 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 3 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 4 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATcttTCCATGGACCCCAATGAGAACCCAC  
> Seq 5 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 6 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 7 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATcttTCCATGGACCCCAATGAGAACCCAC  
> Seq 8 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 9 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 10 GGGCTCACGAGTCCTTATCGTCTGCTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 11 GGGCTCACGAGTCCTTATCGTctccTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 12 GGGCTCACGAGTCCTTATCGTctccTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC  
> Seq 13 GGGCTCACGAGTCCTTATCGTctccTCCGAGATCAGTCAGTATCTTATTCCATGGACCCCAATGAGAACCCAC

120/150

## FIG. 13 (continued)

> Seq 14 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 15 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTtctcggaTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 16 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 17 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 18 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 19 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 20 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 21 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 22 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 23 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 24 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 25 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 26 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 27 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 28 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 29 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTtctTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 30 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 31 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 32 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtctTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 33 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTatgtTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 34 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTtctgTTATTCCATGGACCCCAATGAGAAACCAC  
 > Seq 35 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCtgTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 36 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
 > Seq 37 GGGCTCAGAGTCCTTATCGTCTcctCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC

## FIG. 13 (continued)

> Seq 38 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
> Seq 39 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 40 GGGCTCAGAGTCCTTATCGTCTcTCCGAGATCACTGCTATCggaTTCCATGGACCCCAATGAGAAACCAC  
> Seq 41 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 42 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTcTCCATGGACCCCAATGAGAAACCAC  
> Seq 43 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTcTCCATGGACCCCAATGAGAAACCAC  
> Seq 44 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTcTCCATGGACCCCAATGAGAAACCAC  
> Seq 45 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 46 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 47 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 48 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 49 GGGCTCAGAGTCCTTATCGTCTcTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 50 GGGCTCAGAGTCCTTATCGTCTGcTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 51 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 52 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 53 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 54 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 55 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 56 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 57 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 58 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 59 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC  
> Seq 60 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTcTCCATGGACCCCAATGAGAAACCAC  
> Seq 61 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATCTTATTCCATGGACCCCAATGAGAAACCAC

122/150

## FIG. 13 (continued)

```
> Seq 62 GGGCTCAGAGTCCTTATCGTCTGCTCCGAGATCACTGCTATcttctTCCATGACCCCAATGAGAACCCAC

631 . . . . . 700
> Seq 1 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 2 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 3 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 4 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 5 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 6 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 7 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 8 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 9 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 10 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 11 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 12 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 13 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 14 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 15 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 16 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 17 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 18 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 19 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 20 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
> Seq 21 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGGTTCAGGCCCTC
```

## FIG. 13 (continued)

> Seq 22 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 23 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 24 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 25 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 26 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 27 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 28 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 29 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 30 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 31 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 32 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 33 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 34 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 35 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 36 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 37 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 38 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 39 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 40 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 41 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 42 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 43 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 44 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
> Seq 45 CTTGACTCACTCGTCGCTCAAGCTTTATTCCGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC

124/150

## FIG. 13 (continued)

> Seq 46 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 47 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 48 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 49 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 50 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 51 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 52 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 53 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 54 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 55 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 56 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 57 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 58 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 59 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 60 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 61 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC  
 > Seq 62 CTTGACTCACTCGTCGCTCAAGCTTTATTTCGGAGACGGAGCTGCAGCACTCATTTGTGGTTTCAGGCCCTC

701

> Seq 1 ACTTGCCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAA  
 > Seq 2 ACTTGCCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAA  
 > Seq 3 ACTTGCCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAA  
 > Seq 4 ACTTGCCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAA  
 > Seq 5 ACTTGCCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAACAATCTTGCCGGACACTGAGAA

770

125/150

## FIG. 13 (continued)

> Seq 6 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 7 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 8 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 9 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 10 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 11 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 12 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 13 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 14 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 15 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 16 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 17 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 18 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 19 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 20 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 21 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 22 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 23 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 24 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 25 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 26 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 27 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 28 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 29 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA

126/150

## FIG. 13 (continued)

> Seq 30 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 31 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 32 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 33 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 34 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 35 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 36 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 37 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 38 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 39 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 40 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 41 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 42 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 43 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 44 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 45 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 46 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 47 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 48 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 49 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 50 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 51 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 52 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
> Seq 53 ACTTGGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA

127/150

## FIG. 13 (continued)

> Seq 54 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 55 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 56 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 57 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 58 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 59 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 60 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 61 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA  
 > Seq 62 ACTTGCCGTAGAACGGCCAAATATTCGAGATCGTGTCAACTGATCAAAACAATCTTGCCGGACACTGAGAA

771 . . . . . 840  
 > Seq 1 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 2 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 3 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 4 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 5 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTtcACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 6 GGCAATGAAGTTACACTTGAGAGAGGGAGGGgatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 7 GGCAATGAAGTTACACTTGAGAGAGGGAGGGgatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 8 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 9 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 10 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 11 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 12 GGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
 > Seq 13 GGCActgAAGaacCACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC

## FIG. 13 (continued)

> Seq 14 GGCActgAAGaaccACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 15 GGCActgAAGaaccACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 16 GGCActgAAGaaccACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 17 GGCActgAAGaaccACTTGAGAGAGGGGgGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 18 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 19 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 20 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 21 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 22 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 23 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 24 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 25 GGCAtttAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 26 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 27 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 28 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 29 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 30 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 31 GGCAATGAAGTTACACTTGAGAGAGGGGgGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 32 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 33 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 34 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 35 GGCAATGAAGTTACACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 36 GGCActgAAGaaccACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 37 GGCActgAAGaaccACTTGAGAGAGGGGgTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC

129/150

## FIG. 13 (continued)

> Seq 38 GGCAATGAAGTTACACTTGAGAGAGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 39 GGCActgAAGaaccACTTGAGAGAGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 40 GGCAATGAAGTTACACTTGAGAGAGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 41 GGCAATGAAGTTACACTTGAGAGAGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 42 GGCAATGAAGTTACACTTGAGAGAGGAGGGGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 43 GGCAATGAAGTTACACTTGAGAGAGGAGGGGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 44 GGCAATGAAGTTACACTTGAGAGAGGAGGGGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 45 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 46 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 47 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 48 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 49 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 50 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 51 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 52 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 53 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 54 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 55 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 56 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 57 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 58 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 59 GGCAATGAAGTTACACTTGAGAGAGGAGGGGTTGACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 60 GGCAATGAAGTTACACTTGAGAGAGGAGGGGttcACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC  
> Seq 61 GGCAATGAAGTTACACTTGAGAGAGGAGGGGatgACGTTTCAGTTGCATAGAGATGTACCCCTTGATGGTC

130/150

## FIG. 13 (continued)

> Seq 62 GGCAATGAAGTTACACTTGAGAGAGGGAGGatgACGTTTCAGTTGCATAGAGATGTACCCTTGATGGTC

841 . . . . . 910

> Seq 1 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 2 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 3 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 4 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 5 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 6 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 7 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 8 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 9 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 10 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 11 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 12 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 13 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 14 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 15 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 16 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 17 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 18 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 19 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 20 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

> Seq 21 GCAAAGAACATAGAGAACGCGAGCGGAGAAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACCTCAGTTT

## FIG. 13 (continued)

> Seq 22 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 23 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 24 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 25 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 26 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 27 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 28 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 29 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 30 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 31 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 32 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 33 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 34 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 35 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 36 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 37 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 38 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 39 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 40 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 41 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 42 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 43 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 44 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT  
 > Seq 45 GCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACCTCAGTTT

132/150

## FIG. 13 (continued)

> Seq 46 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 47 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 48 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 49 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 50 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 51 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 52 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 53 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 54 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 55 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 56 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 57 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 58 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 59 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 60 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 61 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT  
 > Seq 62 GCAAAGAACATAGAGAAACGACGGAGAAAGCGTTGTCTCCACTAGGGATAAAGTATTGGAACTCAGTTT

911 . . . . . 980  
 > Seq 1 TCTGGATGGTGCACCCACAGGTGGTCCGAGCCATATTGGACCAGGTGGAGCGGAAACCTAAACCTTAAGGAAGA  
 > Seq 2 TCTGGATGGTGCACCCACAGGTGGTCCGAGCCATATTGGACCAGGTGGAGCGGAAACCTAAACCTTAAGGAAGA  
 > Seq 3 TCTGGATGGTGCACCCACAGGTGGTCCGAGCCATATTGGACCAGGTGGAGCGGAAACCTAAACCTTAAGGAAGA  
 > Seq 4 TCTGGATGGTGCACCCACAGGTGGTCCGAGCCATATTGGACCAGGTGGAGCGGAAACCTAAACCTTAAGGAAGA  
 > Seq 5 TCTGGATGGTGCACCCACAGGTGGTCCGAGCCATATTGGACCAGGTGGAGCGGAAACCTAAACCTTAAGGAAGA

133/150

## FIG. 13 (continued)

> Seq 6 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 7 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 8 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 9 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 10 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 11 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 12 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 13 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 14 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 15 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 16 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 17 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 18 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 19 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 20 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 21 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 22 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 23 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 24 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 25 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 26 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 27 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 28 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA  
> Seq 29 TCTGGATGGTGCACCCAGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGA

## FIG. 13 (continued)

> Seq 30 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 31 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 32 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 33 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 34 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 35 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 36 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 37 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 38 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 39 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 40 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 41 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 42 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 43 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 44 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 45 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 46 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 47 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 48 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 49 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 50 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 51 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 52 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA  
> Seq 53 TCTGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACCTAAACCTTAAGGAAGA

135/150

## FIG. 13 (continued)

|       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |     |    |    |    |    |    |    |    |    |    |    |    |    |    |   |
|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| > Seq | 54 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 55 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 56 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 57 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 58 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 59 | TC | TG | GA | TG | GT | GC | AG | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 60 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 61 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |
| > Seq | 62 | TC | TG | GA | TG | GT | GC | AC | CC | AG | GT | GC | AG | CC | AT | TT | GG | AC | CAG | GT | GG | AG | CG | AA | AA | CT | AA | AC | CT | TA | AG | GA | A |

|       |     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |      |
|-------|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|------|
|       | 981 | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | 1050 |
| > Seq | 1   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 2   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 3   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 4   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 5   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 6   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 7   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 8   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 9   | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | g  | ct | GT | TC | ATC  |
| > Seq | 10  | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | TG | TC | GT | TC | ATC  |
| > Seq | 11  | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | g  | ct | GT | TC | ATC  |
| > Seq | 12  | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | g  | ct | GT | TC | ATC  |
| > Seq | 13  | TA | AG | TT | AA | GG | GT | AG | CA | GG | CA | TG | TC | TT | AG | TC | GA | TA | AC | GG | AA | AA | CC | CT | GA | TT | AG | CG | CT | g  | ct | GT | TC | ATC  |

## FIG. 13 (continued)

> Seq 14 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 15 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 16 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 17 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 18 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 19 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 20 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 21 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 22 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC  
> Seq 23 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 24 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 25 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 26 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 27 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGtltAGCGCTTGTGTTGTTTCATC  
> Seq 28 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGatgAGCGCTTGTGTTGTTTCATC  
> Seq 29 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 30 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 31 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 32 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 33 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 34 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 35 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 36 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
> Seq 37 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAAATACGGAAACCTGATTAGCGCTgctGTGTTGTTTCATC

## FIG. 13 (continued)

> Seq 38 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 39 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 40 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTgtGTGTGTTTCATC  
> Seq 41 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 42 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 43 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGtttAGCGCTTGTGTGTTTCATC  
> Seq 44 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGtttAGCGCTTGTGTGTTTCATC  
> Seq 45 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 46 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 47 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTtctGTGTGTTTCATC  
> Seq 48 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTgtGTGTGTTTCATC  
> Seq 49 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 50 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 51 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 52 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 53 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 54 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 55 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 56 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 57 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 58 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAcacCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 59 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 60 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTTCATC  
> Seq 61 TAAGTTAAGGGCTAGCAGGCATGTGCTTAGTGAATACGGAAACCTGatgAGCGCTTGTGTGTTTCATC

FIG. 13 (continued)

62 TAAGTTAAGGCTAGCAGGCAATGTGCTTAGTGAATACGAAACCTGATTAGCGCTTGTGTTGTTTCATC  
1051 . . . . . 1120  
1 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
2 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
3 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
4 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
5 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
6 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
7 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
8 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
9 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
10 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
11 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
12 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
13 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
14 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
15 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
16 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
17 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
18 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
19 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
20 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
21 ATTGACGAGGTGAGGAAGAGATCTATGGCGGAAGGAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG

## FIG. 13 (continued)

> Seq 22 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 23 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 24 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 25 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 26 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 27 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 28 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 29 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 30 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 31 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 32 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 33 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 34 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 35 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 36 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 37 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 38 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 39 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 40 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 41 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 42 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 43 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 44 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG  
> Seq 45 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGGTGAAGGTTTGGATTCCGGTG

140/150

## FIG. 13 (continued)

```
> Seq 46 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATgccGGTG
> Seq 47 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 48 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 49 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 50 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 51 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 52 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 53 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 54 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 55 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 56 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 57 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 58 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 59 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 60 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 61 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG
> Seq 62 ATTGACGAGGTGAGGAAGAGATCTATGCGGGAAGGAAAGAGTACAACCGTGAAGGTTTGGATTGCGGTG

1121 . . . . . 1190
> Seq 1 TTTGTTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGC
> Seq 2 TTTGTTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGC
> Seq 3 TTTGTTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGC
> Seq 4 TTTGTTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGC
> Seq 5 TTTGTTTGGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCTTACTGCTGC
```

## FIG. 13 (continued)

> Seq 6 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 7 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 8 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 9 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 10 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 11 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 12 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 13 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 14 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 15 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 16 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 17 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 18 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 19 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 20 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 21 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 22 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 23 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 24 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 25 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 26 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 27 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 28 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 29 TTTTGTTTGGATTTGGACCGGATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC

## FIG. 13 (continued)

> Seq 30 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 31 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 32 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 33 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 34 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 35 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 36 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 37 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 38 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 39 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 40 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 41 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 42 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 43 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 44 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 45 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 46 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 47 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 48 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 49 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 50 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 51 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 52 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC  
> Seq 53 TTTTGTTCGATTGGACCGGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCCTCCGCTTACTGCTGC

## FIG. 13 (continued)

> Seq 54 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 55 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 56 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 57 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 58 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 59 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 60 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 61 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC  
> Seq 62 TTTTGTTCGATTGGACCGGTATGACTGTTGAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGC

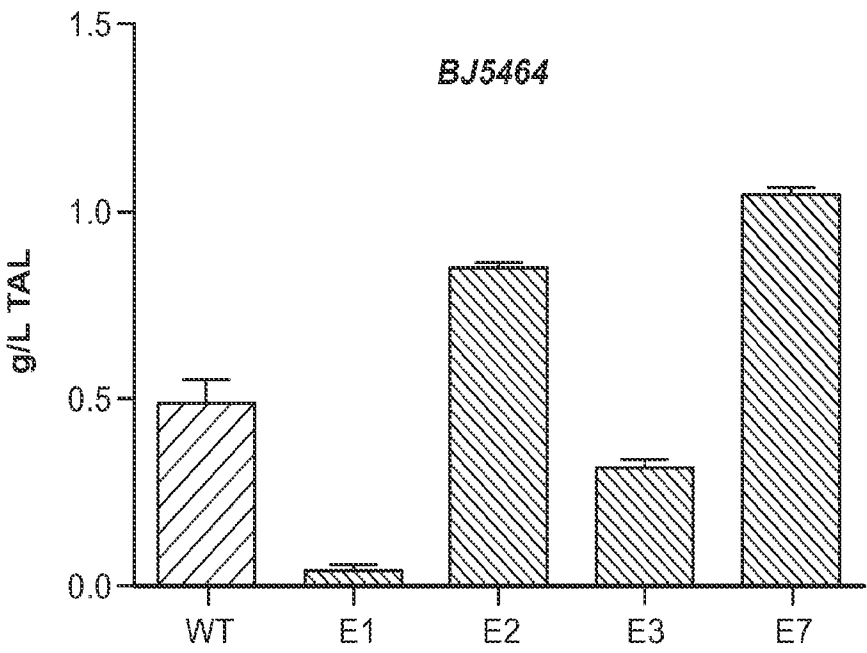
1191 . 1209  
> Seq 1 GGTGCGCAATGGAAACTGA  
> Seq 2 GGTGCGCAATGGAAACTGA  
> Seq 3 GGTGCGCAATGGAAACTGA  
> Seq 4 GGTGCGCAATGGAAACTGA  
> Seq 5 GGTGCGCAATGGAAACTGA  
> Seq 6 GGTGCGCAATGGAAACTGA  
> Seq 7 GGTGCGCAATGGAAACTGA  
> Seq 8 GGTGCGCAATGGAAACTGA  
> Seq 9 GGTGCGCAATGGAAACTGA  
> Seq 10 GGTGCGCAATGGAAACTGA  
> Seq 11 GGTGCGCAATGGAAACTGA  
> Seq 12 GGTGCGCAATGGAAACTGA  
> Seq 13 GGTGCGCAATGGAAACTGA

## FIG. 13 (continued)

> Seq 14 GGTGCCCCAATGGAAACTGA  
> Seq 15 GGTGCCCCAATGGAAACTGA  
> Seq 16 GGTGCCCCAATGGAAACTGA  
> Seq 17 GGTGCCCCAATGGAAACTGA  
> Seq 18 GGTGCCCCAATGGAAACTGA  
> Seq 19 GGTGCCCCAATGGAAACTGA  
> Seq 20 GGTGCCCCAATGGAAACTGA  
> Seq 21 GGTGCCCCAATGGAAACTGA  
> Seq 22 GGTGCCCCAATGGAAACTGA  
> Seq 23 GGTGCCCCAATGGAAACTGA  
> Seq 24 GGTGCCCCAATGGAAACTGA  
> Seq 25 GGTGCCCCAATGGAAACTGA  
> Seq 26 GGTGCCCCAATGGAAACTGA  
> Seq 27 GGTGCCCCAATGGAAACTGA  
> Seq 28 GGTGCCCCAATGGAAACTGA  
> Seq 29 GGTGCCCCAATGGAAACTGA  
> Seq 30 GGTGCCCCAATGGAAACTGA  
> Seq 31 GGTGCCCCAATGGAAACTGA  
> Seq 32 GGTGCCCCAATGGAAACTGA  
> Seq 33 GGTGCCCCAATGGAAACTGA  
> Seq 34 GGTGCCCCAATGGAAACTGA  
> Seq 35 GGTGCCCCAATGGAAACTGA  
> Seq 36 GGTGCCCCAATGGAAACTGA  
> Seq 37 GGTGCCCCAATGGAAACTGA

## FIG. 13 (continued)

> Seq 38 GGTGCCCAATGGAAACTGA  
> Seq 39 GGTGCCCAATGGAAACTGA  
> Seq 40 GGTGCCCAATGGAAACTGA  
> Seq 41 GGTGCCCAATGGAAACTGA  
> Seq 42 GGTGCCCAATGGAAACTGA  
> Seq 43 GGTGCCCAATGGAAACTGA  
> Seq 44 GGTGCCCAATGGAAACTGA  
> Seq 45 GGTGCCCAATGGAAACTGA  
> Seq 46 GGTGCCCAATGGAAACTGA  
> Seq 47 GGTGCCCAATGGAAACTGA  
> Seq 48 GGTGCCCAATGGAAACTGA  
> Seq 49 GGTGCCCAATGGAAACTGA  
> Seq 50 GGTGCCCAATGGAAACTGA  
> Seq 51 GGTGCCCAATGGAAACTGA  
> Seq 52 GGTGCCCAATGGAAACTGA  
> Seq 53 GGTGCCCAATGGAAACTGA  
> Seq 54 GGTGCCCAATGGAAACTGA  
> Seq 55 GGTGCCCAATGGAAACTGA  
> Seq 56 GGTGCCCAATGGAAACTGA  
> Seq 57 GGTGCCCAATGGAAACTGA  
> Seq 58 GGTGCCCAATGGAAACTGA  
> Seq 59 GGTGCCCAATGGAAACTGA  
> Seq 60 GGTGCCCAATGGAAACTGA  
> Seq 61 GGTGCCCAATGGAAACTGA  
> Seq 62 GGTGCCCAATGGAAACTGA



|                  |     |     |   |      |           |
|------------------|-----|-----|---|------|-----------|
| Original Residue | --- | C35 |   | L268 | C35, L268 |
| New Residue      | --- | A   | S | F    | S, F      |

FIG. 14

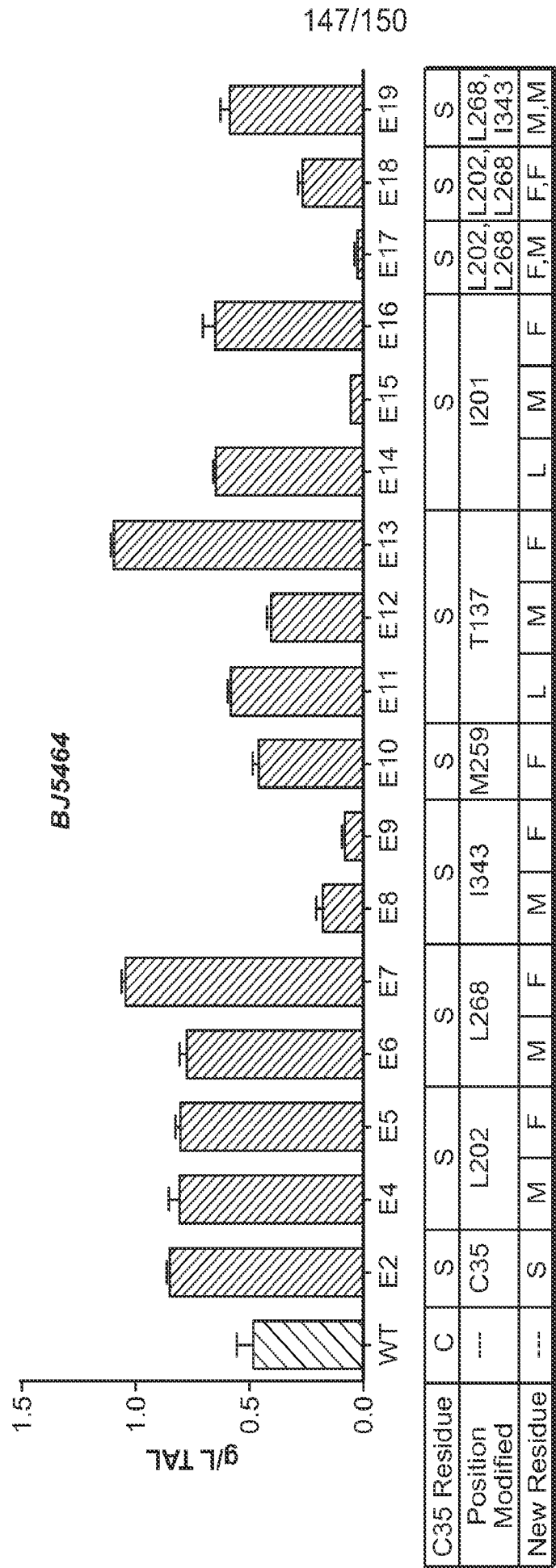


FIG. 15

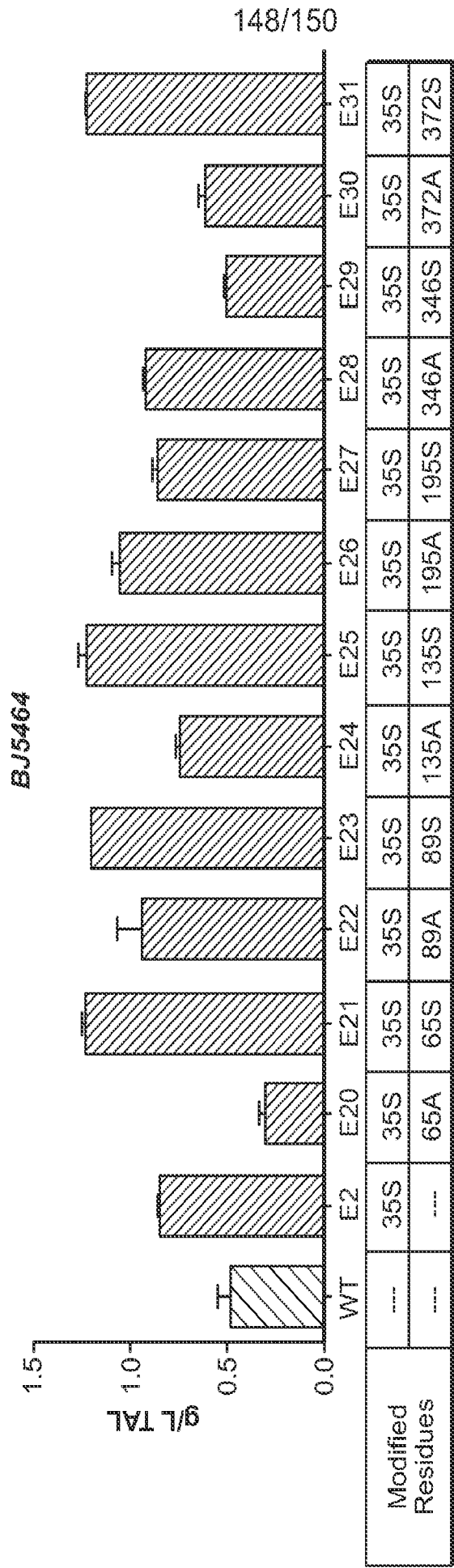


FIG. 16

BJ5464

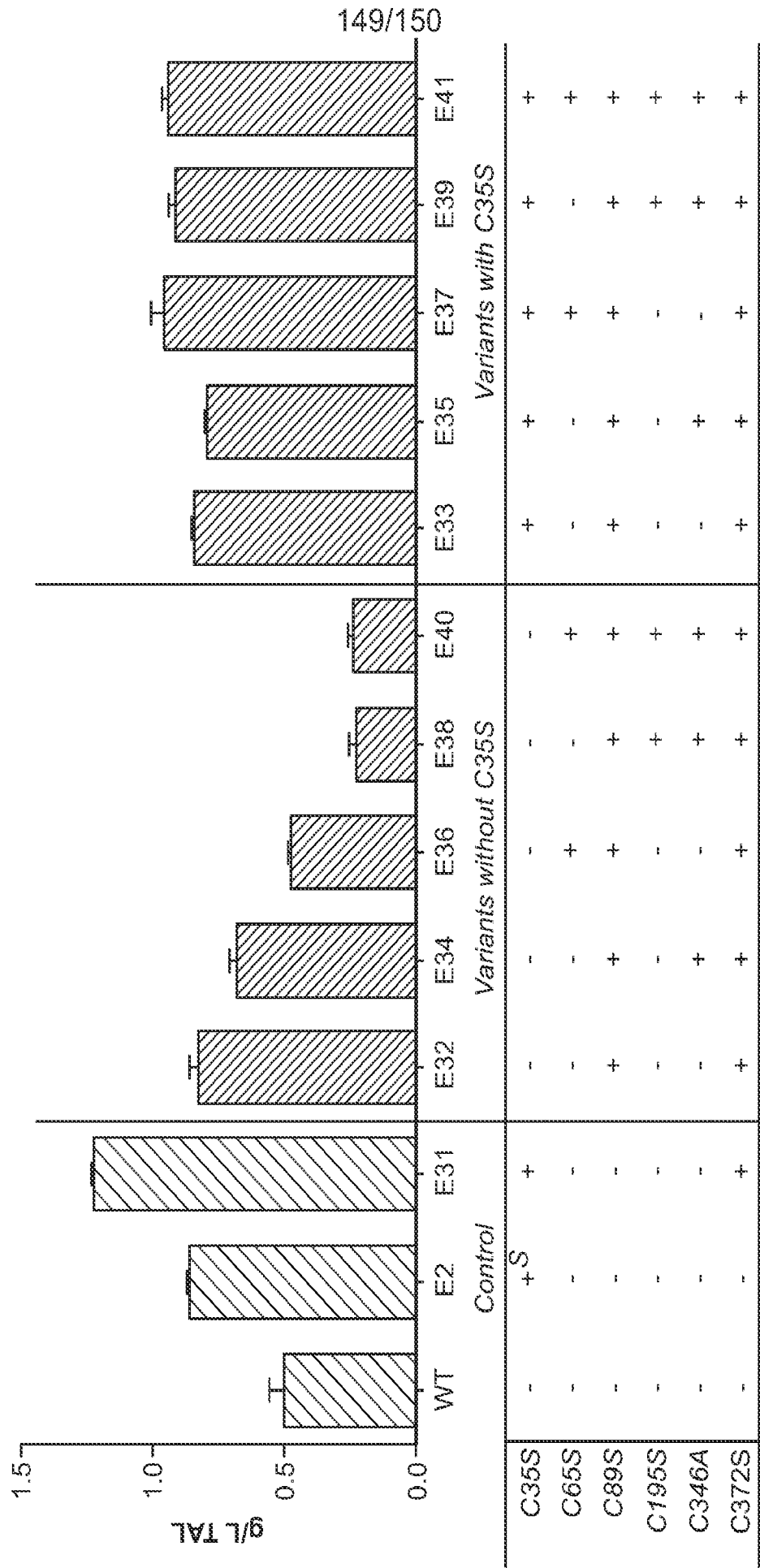


FIG. 17

150/150

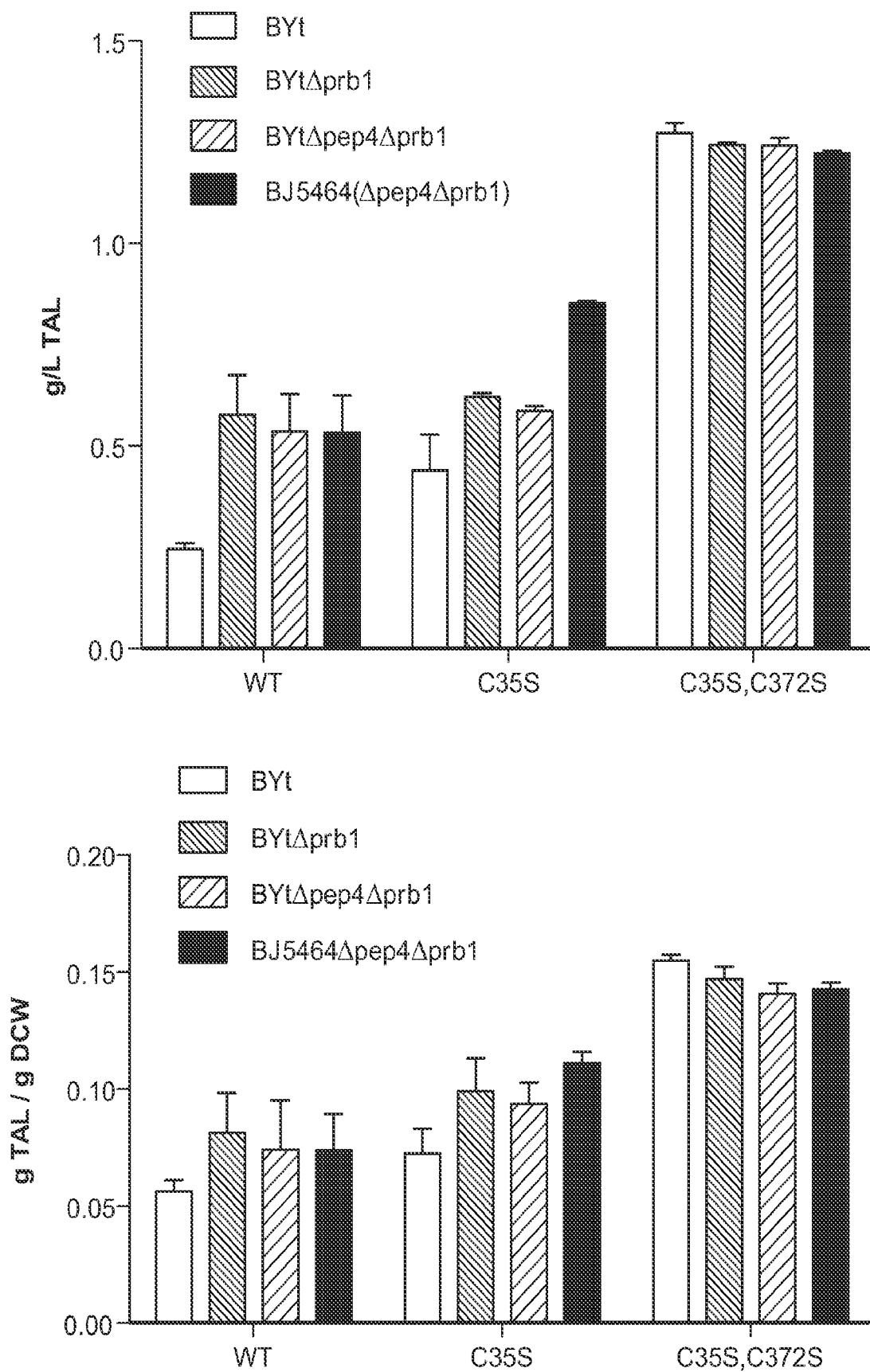


FIG. 18

## INTERNATIONAL SEARCH REPORT

 International application No.  
**PCT/US2015/019058**

## A. CLASSIFICATION OF SUBJECT MATTER

**C12N 9/00 (2006.01) C12N 15/29 (2006.01) A01H 1/00 (2006.01) C07H 21/00 (2006.01) C12N 15/31 (2006.01)**  
**A01H 5/00 (2006.01)**

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Databases used: MEDLINE, AGRICOLA, BIOSIS, HCAPLUS, EMBASE, WPIAP, EPODOC, ESPACENET, GENOMEQUEST, AUSPAT, PUBMED

Keywords used: SALK INSTITUTE FOR BIOLOGICAL STUDIES, UNIVERSITY OF CALIFORNIA, NOEL JOSEPH, BOWMAN MARIANNE, WOODS KATE, POLYKETIDE SYNTHASE, PYRONE SYNTHASE, CHALCONE SYNTHASE, COUMAROYL TRIACETIC SYNTHASE, ACRIDONE SYNTHASE, GH2PS, MUTATE, ALTER, SUBSTITUTE, INCREASE ACTIVITY, ACTIVE SITE, SUBSTRATE SPECIFICITY, SUBSTRATE BINDING, CYSTEINE, ISOLEUCINE, LEUCINE, THREONINE, SEQUENCE P48391, and like terms

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------|-----------------------|
|           | Documents are listed in the continuation of Box C                                  |                       |



Further documents are listed in the continuation of Box C



See patent family annex

|          |                                                                                                                                                                     |     |                                                                                                                                                                                                                                              |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| *<br>"A" | Special categories of cited documents:<br>document defining the general state of the art which is not considered to be of particular relevance                      | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| "E"      | earlier application or patent but published on or after the international filing date                                                                               | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| "L"      | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O"      | document referring to an oral disclosure, use, exhibition or other means                                                                                            | "&" | document member of the same patent family                                                                                                                                                                                                    |
| "P"      | document published prior to the international filing date but later than the priority date claimed                                                                  |     |                                                                                                                                                                                                                                              |

|                                                                                                                                                                |                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Date of the actual completion of the international search<br>5 May 2015                                                                                        | Date of mailing of the international search report<br>05 May 2015                                                                                  |
| <b>Name and mailing address of the ISA/AU</b><br><br>AUSTRALIAN PATENT OFFICE<br>PO BOX 200, WODEN ACT 2606, AUSTRALIA<br>Email address: pct@ipaustalia.gov.au | <b>Authorised officer</b><br><br>Alison Knight<br>AUSTRALIAN PATENT OFFICE<br>(ISO 9001 Quality Certified Service)<br>Telephone No. +61 2 62832847 |

| INTERNATIONAL SEARCH REPORT                           |                                                                                                                                                                                                                                                                                                                                                                                | International application No. |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                | PCT/US2015/019058             |
| Category*                                             | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                             | Relevant to claim No.         |
| X                                                     | TANG SY, et. al. Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter. Journal of the American Chemical Society. 2013, vol. 135, no. 27, pp 10099-103. Figure 2; Tables 2-4; p10099-10100                                                                                            | 1-3; 7-18; 23-28; 30-41       |
| X                                                     | HELARIUTTA Y, et. al. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. Proceedings of the National Academy of Sciences of the United States of America. 1996, vol. 93, no. 17, pp 9033-8. p9033-9034; GenBank Accession numbers X91340 and CAA62683; Figure 3; Abstract | 1; 3; 10-16; 35; 41;          |
| X                                                     | US 2010/0107277 A1 (BRUGLIERA et al) 29 April 2010<br>SEQ ID NO: 28-29; Table 1;                                                                                                                                                                                                                                                                                               | 1; 3; 10-16; 41;              |
| X                                                     | US 2012/0096581 A1 (NOEL et al) 19 April 2012<br>SEQ ID NO: 1; [0195]; Table 1; Fig. 18;                                                                                                                                                                                                                                                                                       | 1; 3; 10-16; 35; 41;          |
| X                                                     | FUNA N, et. al. Alteration of reaction and substrate specificity of a bacterial type III polyketide synthase by site-directed mutagenesis. The Biochemical Journal. 2002, vol. 367, Pt 3, pp 781-9. Figures 1-3; p782-784;                                                                                                                                                     | 1; 3; 10-16; 19; 35-41;       |
|                                                       |                                                                                                                                                                                                                                                                                                                                                                                |                               |

Form PCT/ISA/210 (fifth sheet) (July 2009)

| INTERNATIONAL SEARCH REPORT                                                                                                                                                                                                                                           |                  | International application No. |                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|-------------------------------|------------------|
| Information on patent family members                                                                                                                                                                                                                                  |                  | PCT/US2015/019058             |                  |
| This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information. |                  |                               |                  |
| Patent Document/s Cited in Search Report                                                                                                                                                                                                                              |                  | Patent Family Member/s        |                  |
| Publication Number                                                                                                                                                                                                                                                    | Publication Date | Publication Number            | Publication Date |
| US 2010/0107277 A1                                                                                                                                                                                                                                                    | 29 April 2010    | US 8445748 B2                 | 21 May 2013      |
|                                                                                                                                                                                                                                                                       |                  | AU 2003254410 A1              | 19 Mar 2004      |
|                                                                                                                                                                                                                                                                       |                  | AU 2003254410 B2              | 21 Apr 2011      |
|                                                                                                                                                                                                                                                                       |                  | CN 1678746 A                  | 05 Oct 2005      |
|                                                                                                                                                                                                                                                                       |                  | CN 1678746 B                  | 14 Nov 2012      |
|                                                                                                                                                                                                                                                                       |                  | CO 5721020 A2                 | 31 Jan 2007      |
|                                                                                                                                                                                                                                                                       |                  | EP 1546340 A1                 | 29 Jun 2005      |
|                                                                                                                                                                                                                                                                       |                  | EP 1546340 B1                 | 07 Mar 2012      |
|                                                                                                                                                                                                                                                                       |                  | EP 2292764 A1                 | 09 Mar 2011      |
|                                                                                                                                                                                                                                                                       |                  | EP 2292764 B1                 | 29 May 2013      |
|                                                                                                                                                                                                                                                                       |                  | EP 2292765 A1                 | 09 Mar 2011      |
|                                                                                                                                                                                                                                                                       |                  | EP 2292765 B1                 | 19 Jun 2013      |
|                                                                                                                                                                                                                                                                       |                  | JP 2006512057 A               | 13 Apr 2006      |
|                                                                                                                                                                                                                                                                       |                  | JP 4798574 B2                 | 19 Oct 2011      |
|                                                                                                                                                                                                                                                                       |                  | JP 2010193895 A               | 09 Sep 2010      |
|                                                                                                                                                                                                                                                                       |                  | JP 5002848 B2                 | 15 Aug 2012      |
|                                                                                                                                                                                                                                                                       |                  | US 2007033674 A1              | 08 Feb 2007      |
|                                                                                                                                                                                                                                                                       |                  | US 7612257 B2                 | 03 Nov 2009      |
|                                                                                                                                                                                                                                                                       |                  | WO 2004020637 A1              | 11 Mar 2004      |
| US 2012/0096581 A1                                                                                                                                                                                                                                                    | 19 April 2012    | AU 2002245133 A1              | 30 Jul 2002      |
|                                                                                                                                                                                                                                                                       |                  | US 2004106175 A1              | 03 Jun 2004      |
|                                                                                                                                                                                                                                                                       |                  | US 7285380 B2                 | 23 Oct 2007      |
|                                                                                                                                                                                                                                                                       |                  | US 2009271888 A1              | 29 Oct 2009      |
|                                                                                                                                                                                                                                                                       |                  | US 7807424 B2                 | 05 Oct 2010      |
|                                                                                                                                                                                                                                                                       |                  | WO 02057418 A2                | 25 Jul 2002      |
| End of Annex                                                                                                                                                                                                                                                          |                  |                               |                  |
| Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.                                                                                                                                             |                  |                               |                  |